Linearization Methods in Time Series Analysis

by

Bei Chen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2011

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In this dissertation, we propose a set of computationally efficient methods based on approximating/representing nonlinear processes by linear ones, so-called *linearization*. Firstly, a *linearization* method is introduced for estimating the multiple frequencies in sinusoidal processes. It utilizes a regularized autoregressive (AR) approximation, which can be regarded as a "large p - small n" approach in a time series context. An appealing property of regularized AR is that it avoids a model selection step and allows for an efficient updating of the frequency estimates whenever new observations are obtained. The theoretical analysis shows that the regularized AR frequency estimates are consistent and asymptotically normally distributed. Secondly, a sieve bootstrap scheme is proposed using the linear representation of generalized autoregressive conditional heteroscedastic (GARCH) models to construct prediction intervals (PIs) for the returns and volatilities. Our method is simple, fast and distribution-free, while providing sharp and well-calibrated PIs. A similar linear bootstrap scheme can also be used for diagnostic testing. Thirdly, we introduce a robust lagrange multiplier (LM) test, which utilizes either the bootstrap or permutation procedure to obtain critical values, for detecting GARCH effects. We justify that both bootstrap and permutation LM tests are consistent. Intensive numerical studies indicate that the proposed resampling algorithms significantly improve the size and power of the LM test in both skewed and heavy-tailed processes. Moreover, fourthly, we introduce a nonparametric trend test in the presence of GARCH effects (NT-GARCH) based on heteroscedastic ANOVA. Our empirical evidence show that NT-GARCH can effectively detect non-monotonic trends under GARCH, especially in the presence of irregular seasonal components. We suggest to apply the bootstrap procedure for both selecting the window length and finding critical values. The newly proposed methods are illustrated by applications to astronomical data, to foreign currency exchange rates

as well as to water and air pollution data. Finally, the dissertation is concluded by an outlook on further extensions of *linearization* methods, e.g., in model order selection and change point detection.

Acknowledgements

I would like to express my sincere gratitude and appreciation to both of my supervisors, Dr. Bovas Abraham and Dr. Yulia R. Gel, for their encouragement, guidance and support in every aspect of my study, research and personal life. Their rigorous scholarship and enthusiasm in statistics have inspired and enriched me as a researcher.

Also, I would like to thank my committee members Dr. Mary Thompson, Dr. Tony Wirjanto, Dr. Keith Hipel and Dr. Pierre Duchesne for their invaluable comments and advices which significantly improved this thesis. A very special thanks to Mary, who has always been a role model for me.

I am deeply grateful for the constant help and encouragement from Mary Lou Dufton. Many thanks also go to Dr. Paul Marriott, Dr. Steve Drekic, Dr. Changbao Wu, Lucy Simpson and Marg Feeney.

I am indebted to my dear parents, Volker, Jianzhong, Marguerite and Min, for their unconditional love and support, and for making me strong and courageous along the way. Lastly, I would like to thank my dear husband, Mathieu. Without him, nothing is possible.

To the memory of Omama

To Mathieu

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Real life time series often exhibit nonlinear patterns. In order to capture such nature of observed processes, a large number of models and procedures have been proposed in the literature, e.g., autoregressive conditional heteroscedastic (ARCH)/general ARCH (GARCH) models (Engle, 1982; Bollerslev, 1986), autoregressive fractionally integrated moving average (ARFIMA) model (Granger and Joyeux, 1980; Hosking, 1981) and artificial neural networks (Gately, 1996). However, parameter estimation in the nonlinear scheme can be complicated, typically sensitive to initial values, and therefore not feasible in real-time modeling and forecasting. In practice, one of the widely applied solutions is to approximate/represent a nonlinear process by a linear model, which is referred to as *linearization*, due to its simplicity.

In fact, *linearization* is widely applied in many disciplines of mathematical science. In nonlinear control theory, a fundamental tool for analyzing a nonlinear system is to transform it into a simple linear one by, for example, a Lie bracket (Khalil, 2002); in machine learning, kernel methods are often used to project observations onto a high or even infinite dimensional space where linear procedures can be applied (Hofmann, et al., 2008; Schölkopf et al., 1998); in numerical

analysis, nonlinear integral and ordinary/partial differential equations are usually solved by an orthonormal basis (Darania et al., 2006), e.g., Legendre polynomials; in microeconomics, decision rules are approximated under the state-space approach to *linearization* (Moffatt, 2008) and consequently a unique solution can be obtained. Similarly, in statistics and particularly in time series analysis, the true model of the observed time series is typically unknown or has a complicated structure. A common approach is to approximate the true model by an autoregressive (AR) or a moving average (MA) equation. The applications of AR approximation have a long history that can be traced back to the early papers Akaike (1969, 1974) and Parzen (1974) following the pioneering work of Yule (1921). The AR model has a simple linear structure and its parameter estimation techniques as well as the corresponding properties are well-investigated (see Brockwell and Davis, 2006; Box, et al., 2008, for an overview). One of the most popular methods in frequency estimation is to approximate the generalized spectral density of the periodic process by an AR model (Berk, 1974; Trft and Kumaresan, 1982; Mackisack and Poskitt, 1989, 1990; Li et al., 1994; Quinn and Hannan, 2001; Chen and Gel, 2010). Ray (1993) proposed to fit a "long" AR model to the long memory ARFIMA process for prediction purposes (Ray and Crato, 1996; Poskitt, 2006). This approach avoids the tedious nonlinear optimization procedure for estimating ARFIMA parameters. Also, Politis (2007) demonstrated that the AR approximation is useful in nonlinear financial applications for modeling and forecasting stock market returns. While AR is popular for dealing with data, the MA approximation is a helpful tool in analyzing the structural properties of the underlying process (Bühlmann, 1995). The asymptotics of the nonparametric resampling technique, sieve bootstrap, is investigated based on the MA representation (Kreiss, 1988; Bühlmann, 1997). Another example is that of establishing the mixing properties of various time series via the MA form (Withers, 1981; Doukhan, 1994). In practice, both AR and MA approximations are served as benchmarks for comparison with more complicated nonlinear

models.

In our study, what we mean by *linearization* is either to approximate a nonlinear time series by a linear model or to analyze a nonlinear process in a linear framework. A stationary process $\{Y_t\}_{t\in\mathbb{Z}}$ is called linear if it satisfies

$$Y_t = \sum_{i=-\infty}^{\infty} \phi_i \epsilon_{t-i}, \tag{1.1}$$

where $\sum_{i=-\infty}^{\infty} \phi_i^2 < \infty$ and $\{\epsilon_t\}$ are independently identically distributed (i.i.d.) with $E\epsilon_t = 0$ and $E\epsilon_t^2 < \infty$. A weaker form of equation (1.1) amounts to relaxing the i.i.d. assumption on $\{\epsilon_t\}$ to the assumption of a martingale difference (Hannan, 1973). Note that these assumptions are typically not satisfied for linear representations of nonlinear processes (Francq and Zakoïan, 1998 and 2000).

The charm of linear processes lies in their simplicity. In fact, the totality of their dependence structure is perfectly captured by a single entity, namely the sequence of the coefficients $\{\phi_i\}$, while the estimation procedure of $\{\phi_i\}$ is simple and well-studied. Moreover, the prediction function of linear processes typically achieves optimality, compared to that of a nonlinear setting where finding the best prediction is usually far from trivial (Sugihara and May, 1990; Fan and Yao, 2003). By linearizing the underlying process, we can take advantage of both parametric and nonparametric properties of a linear model and hence simplify as well as hasten the analysis procedure.

In this dissertation, we discuss how *linearization* helps us to analyze nonlinear processes in the following three applications:

1. Estimating multiple frequencies by AR approximation with regularized least squares (LS).

The problem of tracking multiple unknown frequencies is widely encountered in a variety of disciplines, such as astronomy, signal processing and electrical engineering. The unknown frequencies can be obtained by approximating generalized spectral density of a periodic process by an AR model. The advantage is that an AR model has a simple structure and its parameters can be estimated iteratively, which is crucial for online (real-time) applications. Typically, the order of the AR approximation is chosen by information criteria, such as Akaike Information Criterion (AIC) (Akaike, 1974), Bayesian Information Criterion (BIC) (Schwarz, 1978) or corrected AIC (AICc) (Hurvich and Tsai, 1989). However, with an increase of sample size, the model order may change, which leads to re-estimation of all model parameters. We propose a new iterative procedure for frequency detection based on regularizing the empirical information matrix (Chen and Gel, 2010 and 2011). The suggested method avoids the repeated model selection as well as parameter estimation steps, hence optimizes computational costs. Our regularization technique of adding a penalty term to the diagonal of the empirical information matrix is closely related to high dimensional multivariate problems, so-called "large p-small n". While extending the approach of Mackisack and Poskitt (1989, 1990) and Chen and Gel (2010) of tracking a single frequency under the increasing order of an AR approximation, we derive asymptotic properties of the proposed regularized AR (RAR) estimates of multiple frequencies and evaluate performance by numerical examples and a case study on sunspot detection.

2. Constructing prediction intervals (PI) for returns and volatilities in ARCH/GARCH processes by sieve bootstrap (Chen et al., 2011).

The estimation and forecasting of an ARCH/GARCH process is a nonlinear procedure

and hence computationally cumbersome. However, it is well-known that the squared ARCH/GARCH processes can be expressed in a linear AR/ARMA form. Our idea is to employ the linear representation of an ARCH/GARCH equation and then to adapt a sieve bootstrap procedure (Kreiss, 1988; Bühlmann, 1997) to construct the PIs for returns and volatilities. Our method is novel, simple, efficient and distribution-free. The intensive simulation studies indicate that the new re-sampling method provides sharp and well calibrated PIs for both returns and volatilities while reducing computational costs by up to 100 times, compared to other available re-sampling techniques for ARCH/GARCH models. The proposed approach is theoretically justifiable, i.e., the bootstrap PIs are conditionally consistent, given the observed process, and robust to the departure from normality. We illustrate our procedure by an application to Yen/U.S dollar daily exchange rate data.

3. Applications of *linearization* of ARCH/GARCH processes for diagnostic testing of

   (a) ARCH/GARCH effect,

   Detecting ARCH/GARCH effect is a crucial first step in ARCH/GARCH modeling. In the applied literature, the principal tool for detecting ARCH effect is the Lagrange multiplier (LM) test (Breusch and Pagan, 1979; Engle, 1982), which utilizes the linear representation of the squared ARCH/GARCH processes. The parametric LM test based on the asymptotic χ^2 distribution performs poorly when the underlying process is not normal (Gregory, 1989; Demos and Sentana, 1998). We propose to employ the linear bootstrap and the permutation techniques to find the critical value of the LM test. Our approach is consistent, robust to the deviation from normality and hence reliable under various heavy-tailed and skewed distributions.

(b) trend in the presence of ARCH/GARCH effect.

Also, testing for trend in the presence of ARCH/GARCH effect is of substantial practical importance in environmental, financial, medical and social science applications. We introduce a nonparametric trend test based on heteroscedastic Analysis of Variance (ANOVA) with a large number of factor levels (Wang and Van Keilegom, 2007, Wang and Akritas, 2006), the so-called NT-GARCH. The newly proposed test can effectively detect non-monotonic trends under ARCH/GARCH effect, especially when irregular seasonality appears. For small or moderate samples, we suggest to use a bootstrap procedure to obtain the critical values. Our theoretical findings show that the NT-GARCH test statistic is asymptotically normally distributed under the null hypothesis and the bootstrap NT-GARCH test is consistent.

The main contributions of this thesis are to utilize *linearization* in conjunction with bootstrap techniques to solve the estimation, forecasting and diagnostic testing problems in a nonlinear framework. In particular,

1. we utilize the AR(k) approximation with the regularized least squares (RLS) for detection of multiple frequencies for the case of both k and $n \to \infty$, hence, extending the previously obtained results of Mackisack and Poskitt (1989, 1990), Gel and Barabanov (2007) and Chen and Gel (2010). Due to regularization, the repeated model order selection is avoided and the AR parameters are estimated iteratively. Hence, our approach is fast and useful for on-line modeling. We justify the consistency and asymptotic normality of the regularized frequency estimates and illustrate the new tracking procedure by simulation studies and examples on sunspots (Chen and Gel, 2011).

2. we construct PIs for returns and volatilities in ARCH and GARCH processes using the sieve bootstrap (Chen et al., 2011). Adapting the sieve bootstrap procedure in an ARCH and GARCH framework is novel, which substantially decreases computational costs while providing competitively sharp and well calibrated PIs. The proposed method is purely data-driven and sets minimal restrictions on an ARCH/GARCH innovation processes.

3. we introduce the bootstrap and the permutation versions of LM test to detect ARCH/GARCH effect (Chen and Gel, 2011) and nonparametric trend test under ARCH/GARCH innovations (Chen et al., 2011). The proposed tests provide more accurate size and better power compared to the tests depending on the asymptotic distributions. We justify that the bootstrap-based tests are asymptotically consistent. One of the attractive properties of our trend test is that it can effectively detect non-monotonic trends under unidentifiable seasonality with ARCH/GARCH effect. We plan to extend the same bootstrap strategy to online tracking for change-point and regime switching detection.

In the next chapter, we discuss the AR approximation in multiple frequency detection using regularized LS. In Chapter 3, we introduce computationally efficient sieve bootstrap prediction intervals for returns and volatilities in ARCH/GARCH processes. We present the bootstrap/permutation LM test in Chapter 4 and the nonparametric trend test in Chapter 5. Finally, we conclude with summary and future work in Chapter 6.

Chapter 2

Regularized Autoregressive Multiple Frequency Estimation

The problem of tracking unknown frequencies is widely encountered in a variety of applications, ranging from speech recognition in electrical engineering to the study of sunspots in astronomy. Although the topic has been explored for many years (see, for example, Prony, 1795; Pisarenko, 1973; Hannan and Huang, 1993), it continues to attract considerable attention in the statistics and engineering literature (Chen et al., 2000; Quinn and Hannan, 2001; Song and Li, 2006; Duan et al., 2010; Elasmi-Ksibi et al., 2010; Liu et al., 2011). Among various existing methods, the autoregressive (AR) frequency estimation is one of the most popular approaches due to its computational ease and theoretical convenience (Truft and Kumaresan, 1982; Mackisack and Poskitt, 1989, 1990). However, it is well known that the AR-based frequency estimates are asymptotically biased when the order k of an approximating AR model is fixed (Stoica et al., 1987). A simple remedy is to allow the AR order k to increase with sample size n at an appropriate rate. This, however, may lead to deficiency in estimating the covariance matrix. In addition, it implies

that a new AR order needs to be selected upon the arrival of new observations, and consequently all the AR parameters need to be re-estimated.

In order to avoid these shortcomings, Chen and Gel (2010) introduce an alternative approach, so-called regularized AR (RAR) approximation, to detect hidden frequencies. The idea of RAR is to regularize the sample covariance matrix by a nuclear ridge operator, which allows to fit a much "longer" AR model than the one suggested by Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), and hence reduces the approximation bias. Then, the AR parameters are estimated recursively using the regularized least squares (RLS) method (Gel and Fomin, 2001, and Gel and Barabanov, 2007). Note that our technique of regularizing the sample information matrix can be regarded as a "large k – small n" approach in a time series context. In fact, the order of the RAR approximation can be very close to n, so the regularization technique is particularly crucial in this case to avoid deficiency in model identification. With the help of the ridge regularizer, RAR allows to estimate the AR parameters with different level of accuracy, while the number of estimated parameters grows with the sample size. Therefore, the repeated model selection and parameter estimation are avoided as the sample size increases, which makes the RAR procedure especially attractive for online (real-time) modeling.

In practice, as noted in Stoica et al. (1987), we typically encounter spurious frequency estimates for high-order AR approximation. Therefore, we further propose a Robust Trimming Algorithm (RTA) for the RAR frequency estimates. Our numerical studies show that RTA can effectively eliminate spurious roots and outliers, which noticeably increases the accuracy of frequency estimates. Compared to the simulation results of the AR approximation procedure suggested by Mackisack and Poskitt (1989), the RAR frequency estimates with trimming have

smaller mean squared error (MSE)[1], especially when a periodic signal is substantially embedded in noise, i.e. signal to noise ratio (SNR) is low.

In this chapter we generalize the results of Chen and Gel (2010) to a case of tracking multiple frequencies. In particular, we show that the RAR estimates of multiple frequencies are strongly consistent and asymptotically normally distributed. We also illustrate performance of RAR by numerical experiments and a case study on the sunspot data.

2.1 A brief overview of regularization

In statistics, there is typically a trade-off between the bias and the variance of models fit to complex data sets. In order to reduce the bias, it is advantageous to choose a very large number of parameters; however, if the number of parameters is too large relative to the size of the data set, the empirical estimates of the model parameters exhibit high variance and one often faces the problem of overfitting. A fundamental approach to solve this dilemma is regularization. One of the earliest references is Tychonoff (1943), who used regularization in the context of solving integral equations in a numerically stable manner. In the statistics literature, numerous works have been dedicated to regularization, e.g., for model and variable selection, regression, and covariance estimation.

[1] Suppose that $\hat{\omega}$ is an estimate of the true frequency ω, the MSE of $\hat{\omega}$ is defined as

$$\text{MSE}(\hat{\omega}) = \text{Var}(\hat{\omega}) + \Big(\text{Bias}(\hat{\omega}, \omega)\Big)^2. \tag{2.1}$$

In the simulation studies of m Monte Carlo (MC) repetitions, a MC estimate of the MSE is obtained by

$$\hat{\text{MSE}}(\hat{\omega}) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\omega}_i - \frac{1}{m}\sum_{i=1}^{m}\hat{\omega}_i\right)^2 + \frac{1}{m}\sum_{i=1}^{m}\big(\hat{\omega}_i - \omega\big)^2. \tag{2.2}$$

Regularization is particularly important in the context of nonparametric regression. In particular, when the number of predictor variables in the regression model is larger than the sample size, solutions to the LS equations are not unique. Hoerl and Kennard (1970) propose ridge regression to guarantee the uniqueness of the solutions by adding a penalty term to the residual sum of squares. Currently, the counterpart of ridge regression, "lasso" regression is being extensively investigated (Tibshirani, 1996; Meinshausen, 2005; Bunea et al., 2005, 2006; Tibshirani and Taylor, 2010). The "lasso" minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant. Indeed, penalization is not the only form of regularization being used in statistics. For example, in kernel density estimation (Rosenblatt, 1956; Parzen, 1962) the bandwidth of the kernels induces the regularization effects.

Other than in nonparametric regression, regularization is also widely used for the estimation of covariance matrices. Ledoit and Wolf (2003) consider the Steinian shrinkage towards an identity; Wu and Pourahmadi (2003) suggest the Cholesky decomposition of the covariance matrix to bound its inverse from below; Bickel and Levina (2008a, b) introduce regularizing the covariance matrix by thresholding; Bickel and Gel (2011) propose a banding regularization of covariance matrices to estimate parameters and forecast time series. For the proposed RAR method, we add a ridge operator of a nuclear type to regularize the empirical information matrix, which is explained in detail in the following section.

2.2 Regularized AR frequency estimation

Consider a mixed-spectrum process $\{Y_t\}_{t\in\mathbb{Z}}$

$$Y_t = X_t + \epsilon_t \text{ and } X_t = \sum_{j=1}^{q} \rho_j \cos(\omega_j t + \phi_j), \tag{2.3}$$

where ρ_j and ω_j are constants with $\rho_j > 0$ and $0 < \omega_1 < \ldots < \omega_q < \pi$; ϕ_j are independently identically distributed (i.i.d.) random variables uniformly distributed on $[0, 2\pi)$; $\{\epsilon_t\}$ are i.i.d. random variables with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = \sigma^2 < \infty$. Assume that $q \geq 1$ is known, and $\{\epsilon_t\}$ is independent of $\{\phi_j\}$ and hence of $\{X_t\}$. Given observations $\{Y_1, \ldots, Y_n\}$, our goal is to estimate the frequencies

$$\boldsymbol{\omega} = (\omega_1, \ldots, \omega_q)'. \tag{2.4}$$

First, let us review the RAR approach for estimating $\boldsymbol{\omega}$. Consider an AR(k) model

$$a(B)Y_t = \nu_{k,t} \tag{2.5}$$

where B is a backward shift operator ($BY_t = Y_{t-1}$) and $a(z) = 1 + a_1 z + \ldots + a_k z^k$ is a polynomial of degree k. The AR model (2.5) can be written in a state-space form:

$$Y_t = \boldsymbol{\Phi}'_{k,t-1}\boldsymbol{\tau}_k + \nu_{k,t}, \tag{2.6}$$

where $\boldsymbol{\Phi}_{k,t-1} = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k})'$ and $\boldsymbol{\tau}_k = -(a_1, a_2, \ldots, a_k)'$. The RAR frequency estimation procedure consists of the following three steps:

- **Step 1**: Approximate $\{Y_1, \ldots, Y_n\}$ by a "long" AR(k) process, where k increases with sample size n. Note that k may substantially exceed the AR order suggested by AIC and BIC, i.e. $k \gg \log n$.

- **Step 2**: Estimate the vector of unknown AR parameters $\boldsymbol{\tau}_k$ by the iterative RLS method

$$\begin{aligned}
\hat{\boldsymbol{\tau}}_{k,n+1} &= \hat{\boldsymbol{\tau}}_{k,n} + \boldsymbol{\gamma}^{\varepsilon}_{k,n}\boldsymbol{\Phi}_{k,n}(1 + \boldsymbol{\Phi}'_{k,n+1}\boldsymbol{\gamma}^{\varepsilon}_{k,n}\boldsymbol{\Phi}_{k,n+1})^{-1}(Y_{n+1} - \boldsymbol{\Phi}'_{k,n}\hat{\boldsymbol{\tau}}_{k,n}) \\
\boldsymbol{\gamma}^{\varepsilon}_{k,n+1} &= \boldsymbol{\gamma}^{\varepsilon}_{k,n} - \boldsymbol{\gamma}^{\varepsilon}_{k,n}\boldsymbol{\Phi}_{k,n+1}(1 + \boldsymbol{\Phi}'_{k,n+1}\boldsymbol{\gamma}^{\varepsilon}_{k,n}\boldsymbol{\Phi}_{k,n+1})^{-1}\boldsymbol{\Phi}'_{k,n+1}\boldsymbol{\gamma}^{\varepsilon}_{k,n}
\end{aligned} \tag{2.7}$$

with initial conditions $\hat{\boldsymbol{\tau}}_{k,0} = 0$ and $\boldsymbol{\gamma}^{\varepsilon}_{k,0} = (\varepsilon\boldsymbol{\Lambda}_{k,n})^{-1}$. The matrix $\boldsymbol{\gamma}^{\varepsilon}_{k,n}$ is inverse to the

sample information matrix $\hat{\boldsymbol{R}}^{\varepsilon}_{k,n}$, i.e. $\boldsymbol{\gamma}^{\varepsilon}_{k,n} = (\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}$, where

$$\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} = \hat{\boldsymbol{R}}_{k,n} + \varepsilon \boldsymbol{\Lambda}_{k,n} \text{ with } \hat{\boldsymbol{R}}_{k,n} = \sum_{t=1}^{n} \boldsymbol{\Phi}_{k,t} \boldsymbol{\Phi}'_{k,t} \tag{2.8}$$

and $\boldsymbol{\Lambda}_{k,n} = \mathrm{diag}\{\mathrm{e}^{\mu_j}\}_{\mathrm{j}=1}^{\mathrm{k}}$, $\mu_j > 0$, is a ridge regularizer of a nuclear form. Note that $n^{-1}\hat{\boldsymbol{R}}^{\varepsilon}_{k,n}$ is a sample estimate of the covariance matrix $\boldsymbol{R}_k = \{r_{i-j}\}_{i,j=0}^{k-1}$, where $r_{i-j} = E(Y_t Y_{t+i-j})$ is defined as the theoretical autocovariance function (ACVF). For simplicity, the regularizing parameters ε is chosen to be 1 and $(\mu_1, \ldots, \mu_k)'$ are chosen by cross-validation (see discussion in section 2.4).

- **Step 3**: Let $\{\hat{\beta}_j e^{\pm i\hat{\omega}_{k,j}}\}_{j=1}^{q}$ denote the $2q$ roots of $\hat{a}(z)$ which are closest to the unit circle, then the angular positions of these roots

$$\hat{\boldsymbol{\omega}}_k = (\hat{\omega}_{k,1}, \ldots, \hat{\omega}_{k,q})' \tag{2.9}$$

are the estimates of hidden frequencies.

Note that the model order k can be selected a priori to be equal to (or even to exceed) a potential upper bound of all practically fittable AR models, given the current sample size n. Since we use a nuclear form of ridge regularization $\boldsymbol{\Lambda}_{k,n}$, the AR parameters are obtained with different precision, while the number of accurately identified parameters smoothly grows with the sample size. Hence, RAR can be viewed as a smoothed version of model selection.

2.3 Asymptotic properties of RAR estimates

In this section, we extend the results of Mackisack and Poskitt (1989 and 1990) and prove the strong consistency and the asymptotic normality of the RAR frequency estimates $\hat{\boldsymbol{\omega}}_k$. First,

our goal is to show that the RAR frequency estimates $\hat{\boldsymbol{\omega}}_k$ converge almost surely (a.s.) to the vector of unknown frequencies $\boldsymbol{\omega}$. Here we consider a special form of the regularizing parameter $\mu(n) = o(\log n)$. The proof of this result is based on the strict consistency of the RLS estimates of AR parameters. Note that by Theorem 1 of Stoica et al. (1987), we have

$$\boldsymbol{\tau}_k = -\frac{2}{k}\Big(\sum_{j=1}^{q}\cos(\omega_j), \sum_{j=1}^{q}\cos(2\omega_j), \ldots, \sum_{j=1}^{q}\cos(k\omega_j)\Big)' + O\Big(\frac{1}{k^2}\Big). \tag{2.10}$$

Theorem 2.3.1 *Let Y_t be generated by (2.3). Let $\boldsymbol{q}_k = (q_1, \ldots, q_k)' \in \mathbb{R}^k$ denote a $k \times 1$ vector satisfying $\|\boldsymbol{q}_k\| = O(k^{1/2})$. Assume that $\mu(n) = o(\log n)$. If $n \to \infty$ and $k \to \infty$ such that $k^2/n \to 0$, then*

(1) $\big|\boldsymbol{q}_k'(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k)\big| \to 0$ *a.s.*

(2) $\sup_{\theta\in(0,\pi)} \big||\hat{a}(e^{i\theta})|^2 - |a(e^{i\theta})|^2\big| = o(1)$ *a.s.*

(The proof of Theorem 2.3.1 is given in the Appendix.)

Based on the consistency of the RAR parameter estimates $\hat{\boldsymbol{\tau}}_k$ in Theorem 2.3.1, we derive the following almost sure convergence result of $\hat{\boldsymbol{\omega}}_k$.

Theorem 2.3.2 *Under the assumptions of Theorem 2.3.1, $\hat{\boldsymbol{\omega}}_k \to \boldsymbol{\omega}$ a.s., as $n \to \infty$, $k \to \infty$ such that $k^2/n \to 0$.*

(The proof of Theorem 2.3.2 is given in the Appendix.)

Second, we prove the asymptotic normality of the RAR frequency estimates $\hat{\boldsymbol{\omega}}_k$, starting from deriving the asymptotic distribution of $\hat{\boldsymbol{\tau}}_k$.

Theorem 2.3.3 *Under the conditions of Theorem 2.3.1 and $E(\epsilon_t^4) = \kappa\sigma^4 < \infty$, if $k \to \infty$ and $n \to \infty$ such that $k^2/n \to 0$, then $\sqrt{n}(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k)$ converges to a Gaussian process with*

covariance function $\boldsymbol{\Sigma}^{\tau}(i,j) = \lim_{k\to\infty} \Theta_k(i,j)$*, where* $\Theta_k(i,j)$ *is the entry* (i,j) *of the* $k \times k$ *matrix* $(\boldsymbol{R}_k^{-1}\boldsymbol{M}_k\boldsymbol{\Sigma}_k\boldsymbol{M}_k'\boldsymbol{R}_k^{-1})$ *with*

$$\boldsymbol{M}_k = \begin{pmatrix} a_1 & a_2 & \dots & a_k & 0 \\ a_2 & a_3 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ a_{k-1} & a_k & \dots & 0 & 0 \\ a_k & 0 & \dots & 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 0 & a_1 & \dots & 0 & 0 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & a_{k-2} & \dots & 1 & 0 \\ 0 & a_{k-1} & \dots & a_1 & 1 \end{pmatrix}, \tag{2.11}$$

and $\boldsymbol{\Sigma}_k = \{\sigma_{ij}^{\varepsilon}\}_{i,j=0}^{k}$ *where*

$$\sigma_{ij}^{\varepsilon} = \begin{cases} \delta_{i,j}\sigma^4 + \sigma^2 \sum_{s=1}^{q} 2\rho_s^2 \cos(\omega_s i)\cos(\omega_s j), & i,j \neq 0, \\ (\kappa-1)\sigma^4 + \sigma^2 \sum_{s=1}^{q} 2\rho_s^2, & i,j = 0. \end{cases} \tag{2.12}$$

(See the Appendix for the proof of Theorem 2.3.3.)

Using the result of Theorem 2.3.3, we derive the asymptotic normality of $\hat{\boldsymbol{\omega}}_k$. Let $a^*(z)$ be a polynomial of degree k, i.e. $a^*(z) = 1 + a_1^* z + \ldots + a_k^* z^k$ and $\boldsymbol{\tau}_k^* = -(a_1^*, \ldots, a_k^*)$ such that

$$\boldsymbol{\tau}_k^* = \boldsymbol{R}_k^{+}\boldsymbol{r}_k, \tag{2.13}$$

where $\boldsymbol{R}_k^{+}$ denotes the Moore-Penrose pseudoinverse of $\boldsymbol{R}_k$. The results of Stoica et al. (1989) imply that

$$a^*(z) = B^*(z)A(z), \tag{2.14}$$

where $A(z) = \prod_{s=1}^{q}(1 - 2\cos\omega_s z + z^2)$ and $B^*(z)$ is a monic polynomial of degree $(k-2q)$ uniquely defined by

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|B^*(e^{i\omega})\right|^2 \left|A(e^{i\omega})\right|^2 d\omega = \min_{\{B\}} \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|B(e^{i\omega})\right|^2 \left|A(e^{i\omega})\right|^2 d\omega. \tag{2.15}$$

Note that $A(z)$ has q pairs of roots located on the unit circle at $e^{\pm i\omega_s}$, $s = 1, \ldots, q$. The remaining roots of $a^*(z)$, which are the roots of $B^*(z)$, are located outside the unit circle. For the large value of k, the roots of $B^*(z)$ may be located very close to the unit circle, which eventually causes spurious frequency estimates. (We discuss trimming algorithm of such spurious roots in the next section.) The following theorem states the result on asymptotic normality of $\hat{\boldsymbol{\omega}}_k$.

Theorem 2.3.4 *Under the conditions of Theorem 2.3.1, if* $k^2 \geqslant cn^{1-\delta}$, *for* $0 < \delta < 2/3$, *such that* $k^2/n \to 0$, *then* $\sqrt{n}(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega})$ *converges to a Gaussian process with covariance function* $\boldsymbol{\Sigma}^{\boldsymbol{\omega}}(i,j) = \lim_{k\to\infty} \Psi_k(i,j)$, *where* $\Psi_k(i,j)$ *is the entry* (i,j) *of the* $k \times k$ *matrix* $(\boldsymbol{F}\boldsymbol{G}\boldsymbol{R}_k^{-1}\boldsymbol{M}_k\boldsymbol{\Sigma}_k\boldsymbol{M}_k'\boldsymbol{R}_k^{-1}\boldsymbol{G}'\boldsymbol{F}')$ *with*

$$\boldsymbol{F} = \begin{pmatrix} \frac{\psi_1}{(\theta_1^2+\psi_1^2)} & 0 & \ldots & 0 & -\frac{\theta_1}{(\theta_1^2+\psi_1^2)} & 0 & \ldots & 0 \\ 0 & \frac{\psi_2}{(\theta_2^2+\psi_2^2)} & \ldots & 0 & 0 & -\frac{\theta_2}{(\theta_2^2+\psi_2^2)} & \ldots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \ldots & \frac{\psi_q}{(\theta_q^2+\psi_q^2)} & 0 & 0 & \ldots & -\frac{\theta_q}{(\theta_q^2+\psi_q^2)} \end{pmatrix}, \quad (2.16)$$

$\boldsymbol{G} = (\boldsymbol{h}_1', \ldots, \boldsymbol{h}_q', \boldsymbol{g}_1', \ldots, \boldsymbol{g}_q')'$, *for* $s = 1, \ldots, q$,

$$\begin{aligned} &\theta_s = (\cos\omega_s, 2\cos 2\omega_s, \ldots, k\cos k\omega_s)\boldsymbol{\tau}_k^*, \;\; \psi_s = (\sin\omega_s, 2\sin 2\omega_s, \ldots, k\sin k\omega_s)\boldsymbol{\tau}_k^*, \\ &\boldsymbol{h}_s = (\cos\omega_s, \cos 2\omega_s, \ldots, \cos k\omega_s)', \;\; \boldsymbol{g}_s = (\sin\omega_s, \sin 2\omega_s, \ldots, \sin k\omega_s)'. \end{aligned} \quad (2.17)$$

(The proof of Theorem 2.3.4 is given in the Appendix.)

Hence, we conclude that the RAR estimates $\hat{\boldsymbol{\omega}}_k$ of multiple frequencies are strongly consistent and asymptotically normally distributed. In addition, the advantage of RAR in practice is that it avoids the model order selection step and allows to fit a "longer" AR model compared to the one chosen by AIC/BIC, which leads to less bias estimates.

Remark 1. Based on the results of Li et al. (1994), Lau et al. (2002) and the classical results of Bartlett (see Bartlett, 1955; Brockwell and Davis, 1991, Theorem 7.2.1 and Proposition 7.3.1), the asymptotic results of Theorems 2.3.1–2.3.4 can be extended to a more general condition when $\{\epsilon_t\}$ is a linear process of the form

$$\epsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \xi_{t-j}, \tag{2.18}$$

where $\{\xi_t\}$ are i.i.d random variables with $E(\xi_t) = 0$, $E(\xi_t^2) = \sigma_\xi^2$ and $\{\psi_t\}$ is an absolutely summable deterministic sequence with $\sum |\psi_j| < \infty$. In this case, $\{\epsilon_t\}$ is referred to as "colored noise".

Remark 2. Note that here we consider a "soft" version of regularizer $\mu(n) = o(\log n)$, which requires the same condition on k as the unregularized case (Mackisack and Poskitt, 1989 and 1990) to obtain the asymptotic results, i.e., $k^2/n \to 0$. Recalling the discussion of Lau et al. (2002) on the behavior of an AR(k) spectral estimator when both k and $n \to \infty$, an interesting question is: Can we increase the rate of AR approximation k while properly balancing the bias-variance issue with a "stronger" regularizer? I.e., is there theoretical evidence to justify the properties of RAR estimates under $k^{2-\eta}/n \to 0$, for $0 < \eta < 2$? For example, potential regularizers can be the nuclear exponential or polynomial operators with increasing diagonal values, i.e. $\mathbf{\Lambda}_{k,n} = \mathrm{diag}\{\mathrm{e}^{\mu \mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{k}}$ (Gel and Barabanov, 2007) or $\mathbf{\Lambda}_{k,n} = \mathrm{diag}\{\mathrm{j}^{\mathrm{p}}\}_{\mathrm{j}=1}^{\mathrm{k}}$ (Barabanov and Gel, 2005). After numerous unsuccessful attempts to derive the asymptotic properties of the RAR approximation of a higher order k, we decide to leave it as a conjecture:

Conjecture 1 *Let Y_t be generated by (2.3) and let $\mathbf{\Lambda}_{k,n} = \mathrm{diag}\{\mathrm{e}^{\mu \mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{k}}$. As $k \to \infty$, $n \to \infty$ such that $k/n \to 0$, then for any $\delta \in (0,1)$, the following hold in probability:*

(1) $V_{n+1} = n^{-\delta}(\hat{\boldsymbol{\tau}}_{k,n+1} - \boldsymbol{\tau}_k)' \hat{\boldsymbol{R}}_{k,n}^{\varepsilon} (\hat{\boldsymbol{\tau}}_{k,n+1} - \boldsymbol{\tau}_k) \to 0$,

(2) $\hat{\boldsymbol{R}}_{k,n}^{\varepsilon} - C > 0$*,*

which implies that $n^{(1-\delta)/2}\left|(\hat{\boldsymbol{\tau}}_{k,n+1} - \boldsymbol{\tau}_k)\right| \to 0$ *in probability.*

Note that, in the rest of this thesis, a matrix $M > 0$ means that M is positive definite. See Appendix for the proof of condition (2). However, we could not derive condition (1) and hence leave it as an open problem. This conjecture is empirically supported by a number of simulations discussed in the Section 2.5.

2.4 Robust trimming algorithm (RTA)

For sufficiently large n, the roots of $\hat{a}(z)$ are close to those of $a^*(z)$, or equivalently, $B^*(z)A(z)$. As noted in Stoica et al.(1987), the roots of $B^*(z)$ tend to the unit circle when the approximation order k increases. Therefore, the roots of $B^*(z)$ can move faster towards the unit circle than those of $A(z)$, which results in spurious roots and hence leads to false frequency estimates.

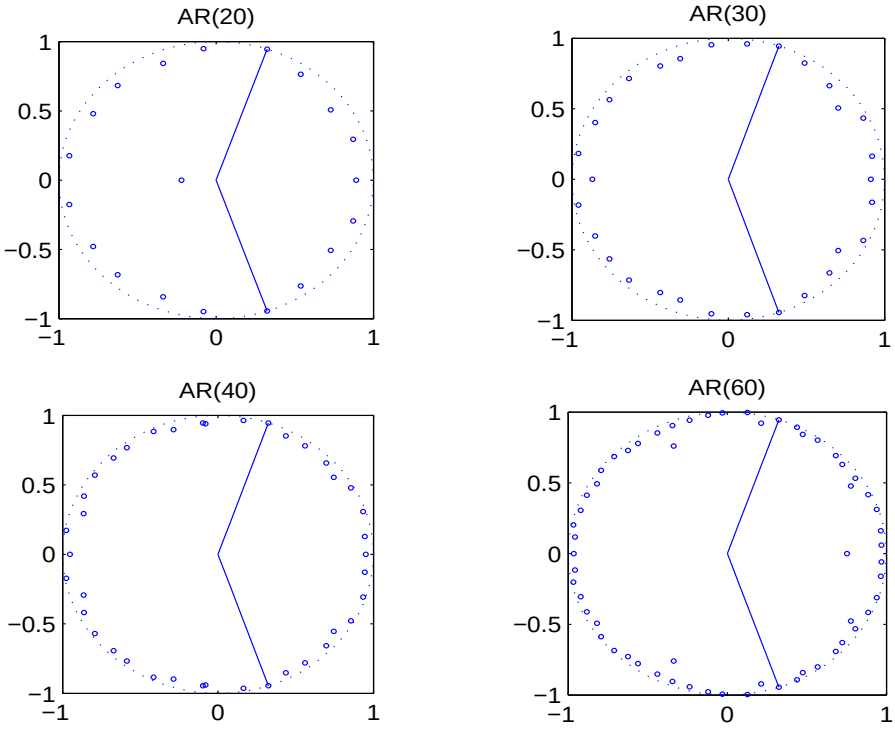

Figure 2.1: Roots of the autoregressive approximating polynomial $\hat{a}(z)$ for different approximation orders $k = 20, 30, 40, 60$. Solid lines are the true angular positions.

Let us consider a sample $\{Y_1, \ldots, Y_{500}\}$ simulated from following process (Mackisack and Poskitt, 1990),

$$Y_t = 20\cos(1.24t + 0.01) + \epsilon_t, \quad \epsilon_t \sim \mathrm{N}(0, 1). \tag{2.19}$$

Figure 2.1 visualizes the effect of an increasing number of spurious roots of the RAR approximating polynomials of degrees 20, 30, 40 and 60, respectively. In order to reduce this effect and increase the accuracy of the frequency estimates, we propose the following robust trimming

algorithm (RTA).

The RTA procedure is implemented via training and execution stages. At the training stage, we select the "optimal" regularizing parameter μ from a range of potential regularizers, using cross-validation on a subset of n_0 observations in such a way that the first $2n_0/3$ observed points are employed for model training and the next $n_0/3$ observations are used for verification. In addition, we construct the $100(1-\alpha)\%$ empirical confidence interval (ECI) of frequency estimates corresponding to various tested regularizers. Heuristically, in view of Theorem 2.3.4, this empirical distribution of frequency estimates is asymptotically normal, because fitting regularizers with different magnitude can be viewed as fitting AR models of different order[2]. At the execution stage, we apply the RAR procedure with the "optimal" regularizer to the entire sample and estimate an unknown frequency. However, only estimates $\hat{\omega}_k$ falling into the pre-chosen $100(1-\alpha)\%$ ECI are taken into account while the rest are disregarded.

The choice of a training set and proportion of data to be employed for cross-validation is typically empirical and data-driven. Currently various rules of thumb are employed in statistical literature on regularization and cross-validation techniques, e.g. $n_0/3$, $n_0/4$, $\log n_0$ etc, where n_0 is the length of a training set, and in many cases the choice of an "optimal" regularizer is relatively insensitive to the proportion of data, selected for cross-validation (see Bickel and Levina, 2008 and references therein). We perform a small simulation study to investigate sensitivity of

[2]Note that to construct the "enforced" $100(1-\alpha)\%$ CI for frequency estimates, here we utilize sample frequency estimates, yielded by RAR with various regularizers at the cross-validation step; these intervals are wider than CIs, constructed using an asymptotic variance, and are more adaptive if the order of approximation is expected to be changed as they do not explicitly depend on one pre-selected approximation order k. However, if the order of autoregressive approximation is not expected to be updated, then the CIs based an asymptotic variance from Theorem 2.3.4 is a more appropriate choice.

the obtained mean squares error and standard errors to a proportion of data, utilized for cross-validation, and our findings indicate that $n_0/3$ and $n_0/4$ provide a very similar performance (see Table 2.1). In contrast, for frequency estimation $\log n_0$ is found to be a very restrictive choice which frequently is numerically unstable with a poor overall performance, and hence, is omitted.

Sample sizes	$n_0/3$	$n_0/4$
	bias (s.e)	bias (s.e)
500	-0.000121 (0.00037)	-0.000129 (0.00031)
1024	-0.000055 (0.00012)	-0.000054 (0.00010)
1900	-0.000025 (0.00005)	-0.000026 (0.00005)

Table 2.1: Observed bias (standard error) of $\hat{\omega}_k$ using various proportions of data allocated for cross-validation.

The selection of the trimming portion α is also empirically based, with a statistical folklore to trim 10% of data points from each tail. Some systematic guides on trimming criteria can be found in the literature on robust statistics (see Andersen, 2008 and references therein), for example, Léger and Romano (1990) propose to determine the amount of trimming by minimizing variance of estimates. In our studies we investigate various trimming amounts, i.e. $\alpha = 5\%$, $\alpha = 10\%$, $\alpha = 25\%$ etc, and find that a relatively minor trimming of 5% from both tails is typically an appropriate choice in the considered examples. Our simulation analysis indicates that RTA can effectively eliminate the spurious roots caused by the high model order, and therefore improve the accuracy of the estimate.

2.5 Numerical examples

In this section, we demonstrate the performance of the RAR frequency estimation with the robust trimming algorithm by simulation studies, using a "stronger" regularizer $\Lambda_{k,n} = \mathrm{diag}\{\mathrm{e}^{\mu \mathrm{j}}\}_{\mathrm{j}=1}^{\mathrm{k}}$ and μ being selected by a cross-validation procedure. We consider two-sinusoid processes with different combinations of amplitudes and frequencies, as shown in Table 2.2 (Stoica et al., 1989) and $\{\epsilon_t\}$ are i.i.d N(0,1).

Case	ρ_1	ρ_2	ω_1	ω_2
1	$10\sqrt{2}$	$10\sqrt{2}$	0.53π	0.23π
2	$10\sqrt{2}$	$10\sqrt{2}$	0.33π	0.23π
3	$10\sqrt{2}$	$10\sqrt{2}$	0.26π	0.23π
4	$\sqrt{2}$	$\sqrt{2}$	0.53π	0.23π
5	$\sqrt{2}$	$\sqrt{2}$	0.33π	0.23π
6	$\sqrt{2}$	$\sqrt{2}$	0.26π	0.23π
7	$\sqrt{20}$	$\sqrt{2}$	0.53π	0.23π
8	$\sqrt{20}$	$\sqrt{2}$	0.33π	0.23π
9	$\sqrt{20}$	$\sqrt{2}$	0.26π	0.23π

Table 2.2: Amplitudes and frequencies of simulation studies.

First, we investigate the variances of $\hat{\omega}_{k,1}$ and $\hat{\omega}_{k,2}$, denoted as $Var(\hat{\omega}_{k,1})$ and $Var(\hat{\omega}_{k,2})$, under different signal-to-noise ratio (SNR) in Cases 1-3. Note that SNR compares the level of a desired signal to the level of a background noise, which is defined as

$$\text{SNR}(j) = 10 \log_{10} \frac{0.5\rho^2}{j\sigma^2} \text{ dB}. \tag{2.20}$$

Here, we take $\rho = 10\sqrt{2}$ and $j = 9.5, 9, 8.5, 8, \ldots, 1, 0.5$. In all considered cases, $Var(\hat{\omega}_{k,1})$ and

$Var(\hat{\omega}_{k,2})$ are compared to the Cramer-Rao Lower Bound (CRLB) (Stoica et al., 1989b) where

$$\text{CRBL}_1 = 24\sigma^2/(\rho_1^2 T^3) \text{ and } \text{CRBL}_2 = 24\sigma^2/(\rho_2^2 T^3). \tag{2.21}$$

Since $\{\epsilon_t\}$ are assumed to be i.i.d N(0,1) in the simulated samples, σ^2 is 1.

Suppose that we consider a sample size n equal to 2000. Based on the first 700 observations, the cross-validation procedure (Chen and Gel, 2010; Bickel and Gel, 2011) selects an "optimal" regularizing parameter $\mu = 0.11$ and AR order $k = 80$. Figure 2.5 shows $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ respectively compared to CRLB while SNR increases from 0.45 to 26.02. Both $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ monotonically decrease as SNR increases and approach CRLB. Also, notice that the differences between the frequencies ω_1 and ω_2 in Case 1, 2 and 3 are correspondingly 0.2π, 0.1π and 0.03π. As the distance between frequencies decreases, the rate of convergence of $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ to CRLB also decreases.

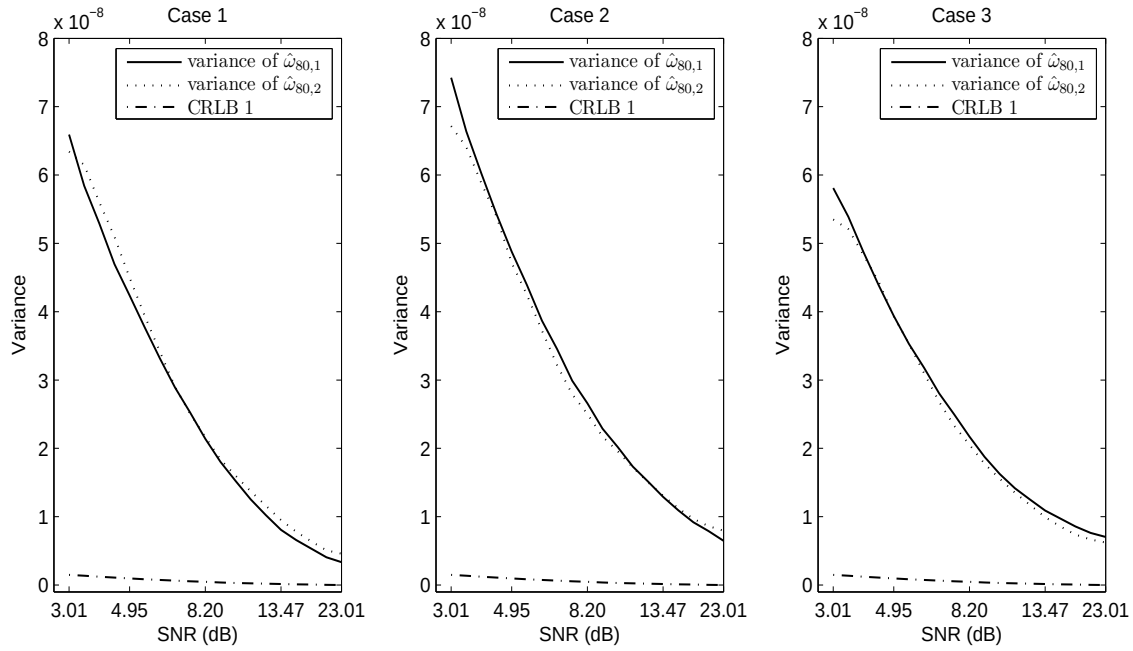

Figure 2.2: The variances of the frequency estimates $\hat{\omega}_{80,1}$ and $\hat{\omega}_{80,2}$ for varying SNR.

Second, we study dynamics of $Var(\hat{\omega}_{k,1})$ and $Var(\hat{\omega}_{k,2})$ with respect to increasing sample size, given SNR of 20dB. Due to the RAR properties, the model order k and regularizing parameter μ remain the same whenever sample size changes, i.e. the previously chosen AR(80) with μ of 0.11 are employed in all cases while T increases from 1000 to 5000. As shown in Figure 2.3-2.5, both $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ strictly decrease as sample size increases. Similar to Figure 2.5, the variances are close to CRLB when the frequencies are well-separated (Case 1, 4 and 7), while the difference becomes larger as the frequencies are closer. Note that the magnitude of variance negatively relates to the amplitude of the sinusoid and hence $Var(\hat{\omega}_{80,1})$ and $Var(\hat{\omega}_{80,2})$ in Figure 2.3 are considerably smaller than those in Figure 2.4 and 2.5.

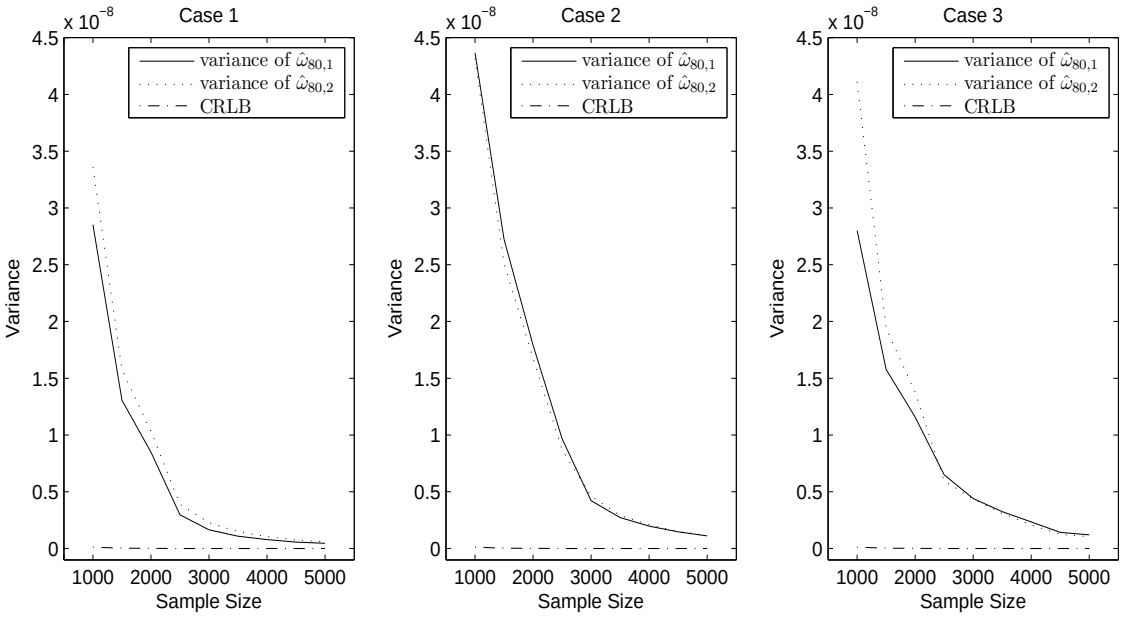

Figure 2.3: The variances of the frequency estimates $\hat{\omega}_{80,1}$ and $\hat{\omega}_{80,2}$ for varying sample sizes when amplitudes $\rho_1 = \rho_2 = 10\sqrt{2}$.

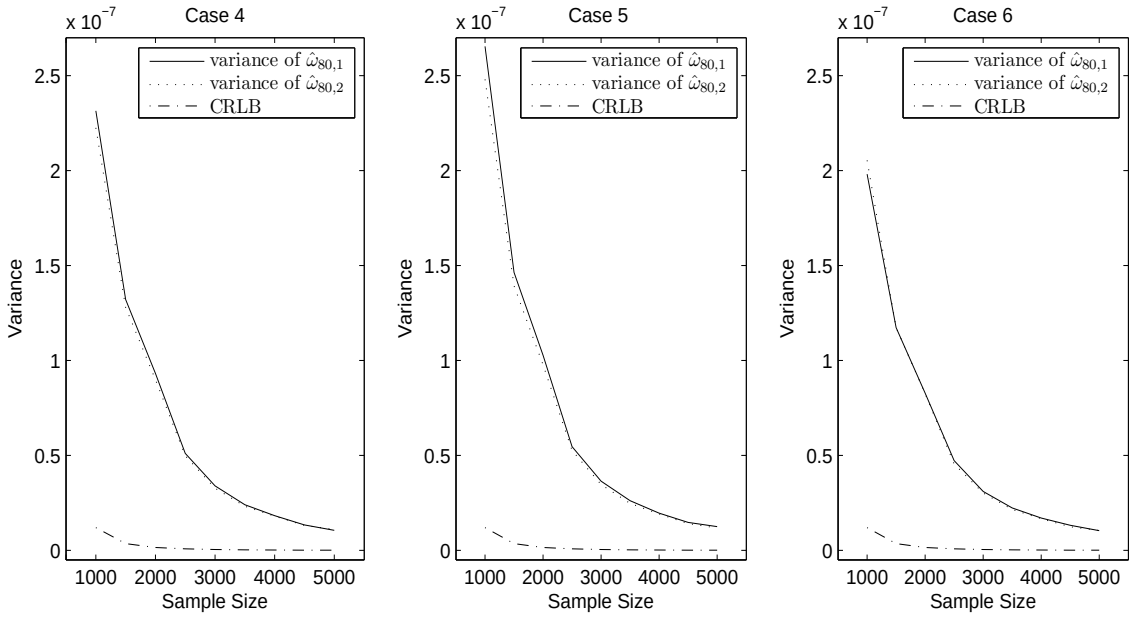

Figure 2.4: The variances of the frequency estimates $\hat{\omega}_{80,1}$ and $\hat{\omega}_{80,2}$ for varying sample sizes when amplitudes $\rho_1 = \rho_2 = \sqrt{2}$.

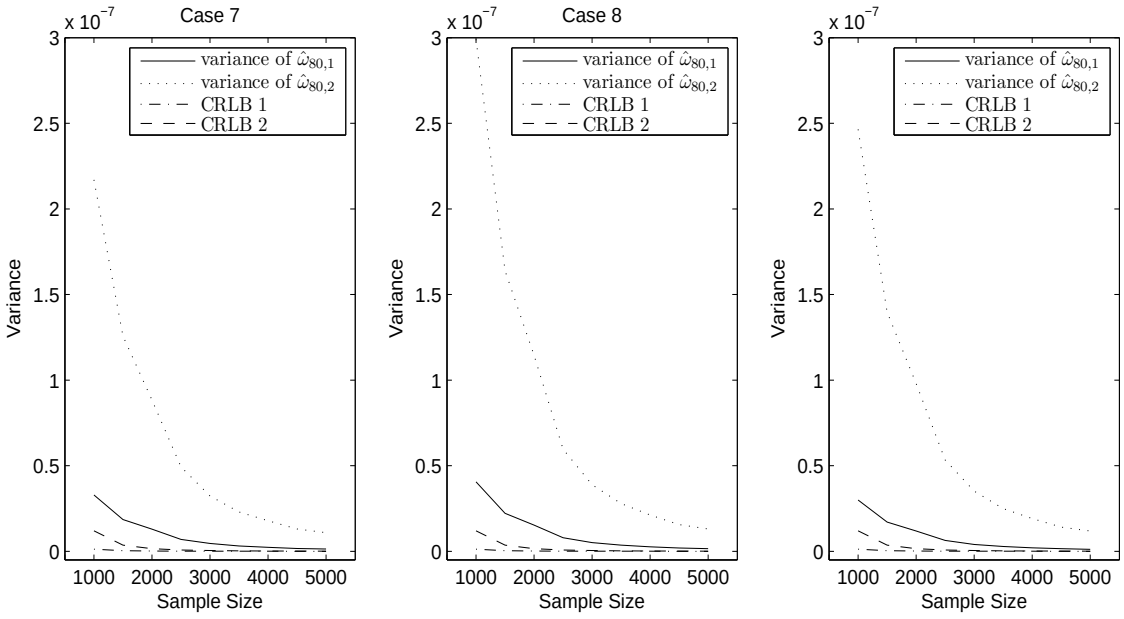

Figure 2.5: The variances of the frequency estimates $\hat{\omega}_{80,1}$ and $\hat{\omega}_{80,2}$ for varying sample sizes when amplitudes $\rho_1 = \sqrt{20}$ and $\rho_2 = \sqrt{2}$.

Since RAR can be viewed as an extension of the results of Mackisack and Poskitt (1989) (from here on referred to as MP), we compare the mean square error (MSE) of RAR to that of MP under varying SNR. Suppose that an observed sample consists of 2000 data points. MP approximates the underlying process by an AR(40) model selected by AIC. Denote a mean squared error (MSE) of the RTA and MP frequency estimates by MSE_{RTA} and MSE_{MP} respectively. Figure 2.7 illustrates the comparison of MSE_{RTA} and MSE_{MP} while SNR increases from 0.45 dB to 26.02 dB. From Figure 2.7, we find that MSE_{RTA} is noticeably smaller than MSE_{MP} when ω_1 and ω_2 are well-separated (Case 1) regardless of SNR; as well as when ω_1 and ω_2 are very close (Case 2) but SNR is low (SNR < 8.43dB). As SNR increases, both MSE_{RTA} and MSE_{MP}

decay exponentially and tend to converge after a certain threshold. Hence, fitting a longer AR model with robust trimming can effectively reduce MSE, especially in noisy conditions which is frequently the case for many applications. From a computational point of view, RTA uses 13.2s CPU time to select the "optimal" regularizer and to construct the 95% CI based on a training set of 700 observations. The MP procedure spends 9.6s CPU time to select an AR order k using AIC for 2000 observations. Although the CPU time for model selection of the MP procedure is somewhat less, the new regularized procedure avoids further model selection and subsequent parameter re-estimation when the sample size increases. [3]

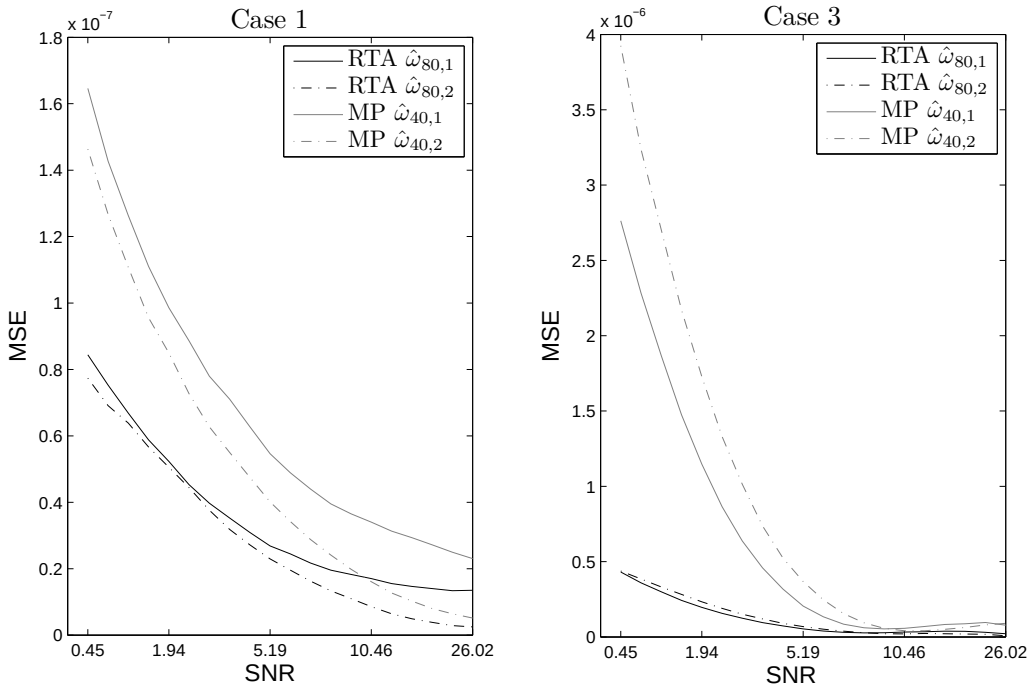

Figure 2.6: Comparison of MSE yielded by MP and RTA for various SNR.

[3]The computations are conducted by a Lenovo Laptop T410 with i7CPU M620 2.67GHz processor.

2.6 Case Study

One of the classical examples of a periodic process is the sunspot observations. The earliest surviving record of sunspot dates from the 364 B.C., according a star catalogue by Chinese astronomer Gan De (Hockey, 1999). In order to demonstrate the proposed RAR method, we take a sample of annual sunspot observations from 1700 to 1988 (see Figure 2.7) and then apply the RAR procedure to estimate the hidden frequency. Using cross-validation, we select an AR(25) model with regularizing parameter $\mu = 0.1$. As a result, the RAR frequency estimate is 0.5721.

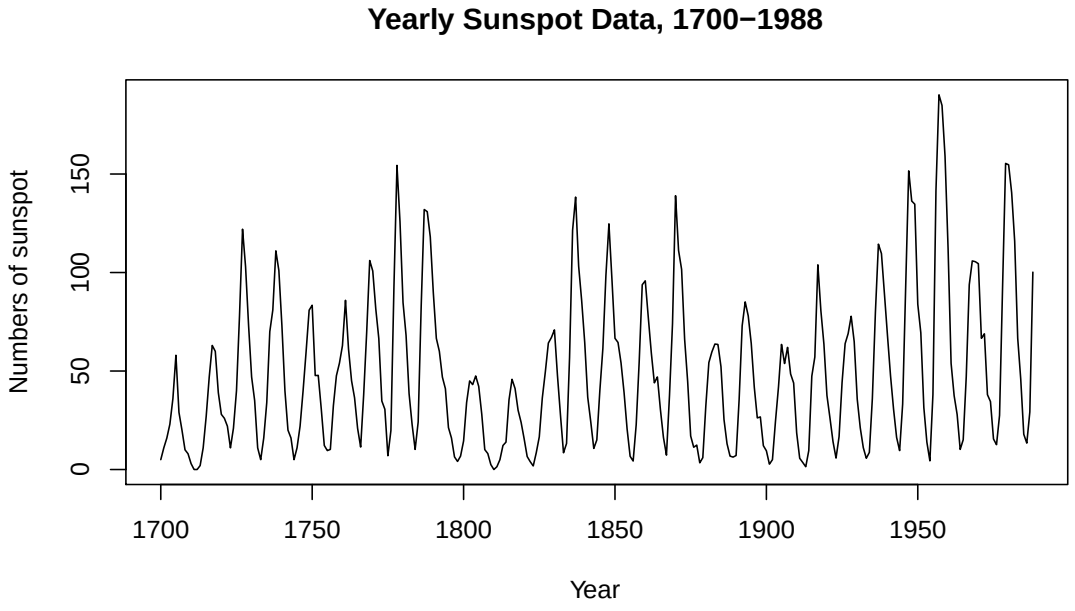

Figure 2.7: Yearly Sunspot data from 1700 to 1988.

We also apply the MP procedure to the sunspot observations for comparison purpose. The AIC selects an AR(9) model and consequently, the MP frequency estimate is 0.3634. In fact, it

is well-known that the sunspot populations rise and fall on an irregular cycle of 11 years, i.e., the hidden frequency is equal to 0.5712. Clearly, the estimation error by RAR is about 0.43% of that by MP.

2.7 Discussion

This chapter generalizes the results of Chen and Gel (2010) on regularized autoregressive (RAR) frequency estimation to a case of multiple unknown periodicities. We show that the RAR estimates of multiple frequencies are strongly consistent and asymptotically normally distributed. Since the idea of RAR is to approximate generalized spectral density of an periodic process by the continuous spectral density of a "long" autoregressive model whose order is substantially higher than the one suggested by AIC or BIC, we encounter a "large k-small n" problem in a time series context. We approach this problem by a nuclear-type ridge regularization of the sample autocovariance matrix and choose an "optimal" regularizer with cross-validation (Chen and Gel, 2010; Bickel and Gel, 2011). Our simulation results indicate that as sample size and/or signal-to-noise ratio increases, the RAR frequency estimates approach the Cramer-Rao Lower Bound, and convergence rate is faster if frequencies are farther apart. Since RAR enables us to avoid frequent re-estimation of approximating model order and parameters, the new procedure is relatively computationally inexpensive and hence feasible for online tracking of unknown multiple frequencies.

The proposed method can be extended to the case of a mixed spectrum process with colored noise, which is a more realistic assumption for a number of applications, e.g. astronomy and speech recognition. Another interesting future extension consists of using banding and thresholding as regularization techniques as well as exploring bootstrap-based selection of an "optimal"

regularizer.

Appendix of Chapter 2

Let us denote covariance vectors by $\boldsymbol{r}_k = (r_1, \ldots, r_k)'$, $\boldsymbol{r}_{k,0} = (r_0, r_1, \ldots, r_k)'$. Also, denote sample ACVF by $\hat{r}_j = \frac{1}{n}\sum_{t=1}^{n-j} Y_t Y_{t+j}$, $j = 0, 1, \ldots, k$, for $k = 0, 1, ..., n-1$, which forms sample covariance vectors $\hat{\boldsymbol{r}}_k = (\hat{r}_1, \ldots, \hat{r}_k)'$ and $\hat{\boldsymbol{r}}_{k,0} = (\hat{r}_0, \hat{r}_1, \ldots, \hat{r}_k)'$.

Proof of Theorem 2.3.1 Let $\boldsymbol{\Gamma}_k$, $\hat{\boldsymbol{\Gamma}}_{k,n}$, $\hat{\boldsymbol{\Gamma}}^{\varepsilon}_{k,n}$ and $\boldsymbol{\Lambda}^{(1)}_{k,n}$ denote the $k \times (k+1)$ matrices formed respectively from the $(k+1) \times (k+1)$ matrices $\boldsymbol{R}_k$, $\hat{\boldsymbol{R}}_{k,n}$, $\hat{\boldsymbol{R}}^{\varepsilon}_{k,n}$ and $\boldsymbol{\Lambda}_{k,n}$ by deleting their first rows. Following the proof of Theorem 3 in Mackisack and Poskitt (1990), we express $(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k)$ as

$$\begin{aligned}
(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) &= \boldsymbol{\gamma}^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{\Gamma}}^{\varepsilon}_{k} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k)' \\
&= \boldsymbol{\gamma}^{\varepsilon}_{k,n}\left(\hat{\boldsymbol{\Gamma}}_{k,n} + \frac{\varepsilon\boldsymbol{\Lambda}^{(1)}_{k,n}}{n} - \boldsymbol{\Gamma}_k\right)(1 : \boldsymbol{\tau}'_k)' \\
&= \boldsymbol{\gamma}^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k)' + \boldsymbol{\gamma}^{\varepsilon}_{k,n}\frac{\varepsilon\boldsymbol{\Lambda}^{(1)}_{k,n}}{n}(1 : \boldsymbol{\tau}'_k)'. \qquad (2.22)
\end{aligned}$$

Since as $n \to \infty$, $n^{-1}\varepsilon e^{\mu(n)} \to 0$ and $\hat{\boldsymbol{R}}_{k,n} \to \boldsymbol{R}_k$ a.s., by Theorem 4.1 of Houdré and Kedem (1995), we obtain

$$\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} = \boldsymbol{R}_k + \big(\hat{\boldsymbol{R}}_{k,n} - \boldsymbol{R}_k\big) + \frac{\varepsilon\boldsymbol{\Lambda}_{k,n}}{n} > 0, \qquad (2.23)$$

and hence $\|\boldsymbol{\gamma}^{\varepsilon}_{k,n}\| \leq C_1$, $C_1 \in \mathbb{R}^+$. Also, by equation (2.10), $\|(1 : \boldsymbol{\tau}'_k)'\| \leq C_2$, $C_2 \in \mathbb{R}^+$, thus $\boldsymbol{\gamma}^{\varepsilon}_{k,n} n^{-1}\varepsilon\boldsymbol{\Lambda}^{(1)}_{k,n}(1 : \boldsymbol{\tau}'_k)' \to 0$ and consequently,

$$(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) \sim \boldsymbol{\gamma}^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k)', \qquad (2.24)$$

which can be re-written as

$$(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) \sim \Big(\boldsymbol{\gamma}^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} - \boldsymbol{R}_k\big) + \boldsymbol{I}_k\Big)\boldsymbol{R}_k^{-1}\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k). \tag{2.25}$$

Equation 2.23 implies that $\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} \to \boldsymbol{R}_k$ a.s. again by Theorem 4.1 of Houdré and Kedem (1995). Following An el al.'s approach (1982, pp.929-930), we can show that

$$\Big(\boldsymbol{\gamma}^{\varepsilon}_{k,n}\big(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} - \boldsymbol{R}_k\big) + \boldsymbol{I}_k\Big) \to \boldsymbol{I}_k, \quad \text{a.s.} \tag{2.26}$$

and hence our problem is reduced to verify

$$\boldsymbol{q}'_k\boldsymbol{R}_k^{-1}\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k) \to 0, \quad \text{a.s.} \tag{2.27}$$

Since $\|\boldsymbol{R}_k^{-1}\| \leq C_3$, $C_3 \in \mathbb{R}^+$, we obtain

$$\|\boldsymbol{q}'_k\boldsymbol{R}_k^{-1}\| = O(\|\boldsymbol{q}_k\|) = O(k^{1/2}). \tag{2.28}$$

Thus, in the rest of the proof we investigate the asymptotic behavior of

$$M_{k,n} = Q'_k\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)(1 : \boldsymbol{\tau}'_k), \tag{2.29}$$

where $Q'_k = \boldsymbol{q}'_k\boldsymbol{R}_k^{-1} = (Q_1, \ldots, Q_k)$.

First, let us consider an element of $\big(\hat{\boldsymbol{\Gamma}}_{k,n} - \boldsymbol{\Gamma}_k\big)$, i.e, $\hat{r}_j - r_j$, $j = 0, \ldots, k$. Denote $r_j^x = E(X_tX_{t+j})$ and $\hat{r}_j^x = \frac{1}{n}\sum_{t=1}^{n-j} X_tX_{t+j}$, we have

$$r_j^x = E\bigg\{\sum_{n=1}^{q}\rho_n\cos(\omega_n t + \phi_n)\sum_{n=1}^{q}\rho_n\cos(\omega_n(t+j) + \phi_n)\bigg\} = \sum_{n=1}^{q}\frac{\rho_n^2}{2}\cos(\omega_n j). \tag{2.30}$$

and thus

$$\begin{aligned}
\hat{r}_j^x - r_j^x &= \frac{1}{n}\sum_{t=1}^{n-j}\sum_{s=1}^{q}\sum_{m=1}^{q}\rho_s\rho_m\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big) - \sum_{s=1}^{q}\frac{\rho_s^2}{2}\cos(\omega_s j) \\
&= \frac{1}{2n}\sum_{s=1}^{q}\rho_s^2\sum_{t=1}^{n-j}\cos\big(\omega_s(2t+j)+2\phi_s\big) \\
&+ \frac{1}{n}\sum_{\substack{s,m=1\\ s\neq m}}^{q}\rho_s\rho_m\sum_{t=1}^{n-j}\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big) \\
&= \frac{1}{2n}\sum_{s=1}^{q}\rho_s^2\sum_{t=1}^{n}\cos\big(\omega_s(2t+j)+2\phi_s\big) - \frac{1}{2n}\sum_{s=1}^{q}\rho_s^2\sum_{t=n-j+1}^{n}\cos\big(\omega_s(2t+j)+2\phi_s\big) \\
&+ \frac{1}{n}\sum_{\substack{s,m=1\\ s\neq m}}^{q}\rho_s\rho_m\sum_{t=1}^{n}\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big) \\
&- \frac{1}{n}\sum_{\substack{s,m=1\\ s\neq m}}^{q}\rho_s\rho_m\sum_{t=n-j+1}^{n}\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big) \qquad (2.31)
\end{aligned}$$

Using the complex exponential representation of cosine function, it can be shown that

$$\left|\sum_{t=1}^{n}\cos\big(\omega_s(2t+j)+2\phi_s\big)\right| \leq \frac{1}{|\sin\omega_s|}, \qquad (2.32)$$

and

$$\left|\sum_{t=1}^{n}\cos(\omega_s t+\phi_s)\cos\big(\omega_m(t+j)+\phi_m\big)\right| \leq \frac{1/2}{\big|\sin\big((\omega_s+\omega_m)/2\big)\big|} + \frac{1/2}{\big|\sin\big((\omega_s-\omega_m)/2\big)\big|} \qquad (2.33)$$

for any ϕ_s, ϕ_m, j and $s \neq m$. Since $\omega_s \in (0,\pi)$ for all s and $\omega_s \neq \omega_m$ for all $s \neq m$, both $1/|\sin\omega_s|$ and $1/\big|\sin\big((\omega_s \pm \omega_m)/2\big)\big|$ can be bounded above by a constant. Therefore, equation (2.31) becomes

$$\hat{r}_j^x - r_j^x = O(1/n) + O(j/n) + O(1/n) + O(j/t) = O(j/n), \qquad (2.34)$$

and hence, we obtain for $j = 0, \ldots, k$:

$$\begin{aligned}
\hat{r}_j - r_j &= \hat{r}_j^x - r_j^x + \frac{1}{n}\sum_{t=1}^{n-j}(x_{t+j}\epsilon_t + x_t\epsilon_{t+j} + \epsilon_t\epsilon_{t+j}) - E(\epsilon_t\epsilon_{t+j}) \\
&= O(j/n) + \frac{1}{n}\sum_{t=1}^{n-j}\sum_{s=1}^{q}\rho_s \cos\big(\omega_s(t+j) + \phi_s\big)\epsilon_t \\
&+ \frac{1}{n}\sum_{t=1}^{n-j}\sum_{s=1}^{q}\rho_s \cos(\omega_s t + \phi_s)\epsilon_{t+j} + \frac{1}{n}\sum_{t=1}^{n-j}\epsilon_t\epsilon_{t+j} - \delta_{j,0}\sigma^2 \\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{q}2\rho_s \cos(\omega_s t + \phi_s)\cos(\omega_s j)\epsilon_t + \frac{1}{n}\sum_{t=1}^{n}(\epsilon_t\epsilon_{t-j} - \delta_{j,0}\sigma^2) \\
&- \frac{1}{n}\sum_{t=1}^{j}\sum_{s=1}^{q}\rho_s \cos\big(\omega_s(t-j) + \phi_s\big)\epsilon_t - \frac{1}{n}\sum_{t=n-j+1}^{n}\sum_{s=1}^{q}\rho_s \cos\big(\omega_s(t+j) + \phi_s\big)\epsilon_t \\
&- \frac{1}{n}\sum_{t=-j+1}^{0}\epsilon_t\epsilon_{t+j} + O\big(\frac{j}{N}\big), \qquad (2.35)
\end{aligned}$$

which implies

$$\begin{aligned}
\big|(\hat{r}_j - r_j) - S_{j,n}\big| &\leqslant \frac{1}{n}\sum_{t=1}^{j}\sum_{s=1}^{q}\rho_s\big|\epsilon_t\big| + \frac{1}{n}\sum_{t=n-j+1}^{n}\sum_{s=1}^{q}\rho_s\big|\epsilon_t\big| \\
&+ \frac{1}{n}\sum_{t=-j+1}^{0}\sum_{s=1}^{q}\rho_s\big|\epsilon_t\epsilon_{t+j}\big| + O\bigg(\frac{j}{n}\bigg), \qquad (2.36)
\end{aligned}$$

where

$$S_{j,n} = \frac{1}{n}\sum_{t=1}^{n}\bigg(\sum_{s=1}^{q}2\rho_s \cos(\omega_s t + \phi_s)\cos(\omega_s j) + \epsilon_{t-j}\bigg)\epsilon_t - \delta_{j,0}\sigma^2. \qquad (2.37)$$

Since ρ_s, $s = 1, \ldots, q$, are constants and $\{\epsilon_t\}$ is assumed to be white noise with finite fourth moment, the four terms on the right-hand side of (2.36) are all $O(j/n)$ a.s. Therefore, for $j = 0, \ldots, k$,

$$\hat{r}_j - r_j = S_{j,n} + O(j/n). \qquad (2.38)$$

Replacing $\hat{r}_j - r_j$ by $S_{j,n} + O(j/n)$ in equation (2.29), we obtain

$$M_{k,n} = Q_k'\big(\boldsymbol{S}_n + \boldsymbol{E}_n\big)(1 : \boldsymbol{\tau}_k')', \tag{2.39}$$

where the matrices $\boldsymbol{S}_n$ and $\boldsymbol{E}_n$ respectively have elements $S_{j-l,n}$ and $O\big((j-l)/n\big)$, $j = 1, \ldots, k$ and $l = 1, \ldots, k+1$, and

$$\big|M_{k,n}\big| \leq \big|Q_k'\boldsymbol{S}_n(1 : \boldsymbol{\tau}_k')'\big| + \big|Q_k'\boldsymbol{E}_n(1 : \boldsymbol{\tau}_k')'\big|. \tag{2.40}$$

As $n \to \infty$ and $k \to \infty$ such that $k^2/n \to 0$, by the Cauchy-Schwartz inequality, we have

$$\big|Q_k'\boldsymbol{E}_n(1 : \boldsymbol{\tau}_k')'\big| \leq \big(O(k)O(k^3/n^2)\big)^{1/2} = O(k^2/n) = o(1). \tag{2.41}$$

Also, note that

$$\big|Q_k'\boldsymbol{S}_n(1 : \boldsymbol{\tau}_k')'\big| \leq \left|\sum_{j=1}^{k} Q_j S_{j,n}\right| + \big|Q_k'\boldsymbol{S}_n(0 : \boldsymbol{\tau}_k')'\big| = \left|\sum_{j=1}^{k} Q_j S_{j,n}\right| + O(\|\boldsymbol{S}_n\|), \tag{2.42}$$

thus it is sufficient to show that $O(\|\boldsymbol{S}_n\|) = o(1)$. By definition,

$$TS_{j,n} = \sum_{m=1}^{b} X_{j,m}, \tag{2.43}$$

where

$$X_{j,m} = \epsilon_m \sum_{s=1}^{q} 2\rho_s \cos(\omega_s m + \phi_s)\cos(\omega_s j) + \epsilon_m\epsilon_{m-j} - \delta_{j,0}\sigma^2. \tag{2.44}$$

The rest of the proof is same as that of Theorem 3 of Mackisack and Poskitt (1990) and hence omit here. □

Proof of Theorem 2.3.2 Let $\boldsymbol{\omega}_k = (\omega_{k,1}, \ldots, \omega_{k,q})'$ be the unknown frequencies based on the k-th order RAR approximation. Note that

$$\hat{\omega}_{k,j} - \omega_j = (\hat{\omega}_{k,j} - \omega_{k,j}) + (\omega_{k,j} - \omega_j), \tag{2.45}$$

for $j = 1, \ldots, q$. Using the result of Theorem 2.3.1 and applying similar arguments as the proofs of Theorem 1 in MacKisack and Poskitt (1989) to a multi-frequency case, we can show that when $n \to \infty$ and $k \to \infty$ such that $k^2/n \to 0$, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} P\Big(\big|\hat{\omega}_{k,j} - \omega_{k,j}\big| \geq \varepsilon\Big) = 0. \tag{2.46}$$

As shown by Stoica et al. (1987), $(\omega_{k,j} - \omega_j) = O(1/k^3)$ and the result follows. □

Denote the regularized sample ACVF by $\hat{r}_j^{\varepsilon} = \frac{1}{n}\big\{\sum_{t=1}^{n-j} Y_t Y_{t+j} + \delta_{j,0}\varepsilon e^{\mu(n)}\big\}$, $j = 0, 1, \ldots, k$, for $k = 0, 1, ..., n-1$, which forms the regularized sample covariance vector $\hat{\boldsymbol{r}}_{k,0}^{\varepsilon} = (\hat{r}_0^{\varepsilon}, \hat{r}_1^{\varepsilon}, \ldots, \hat{r}_k^{\varepsilon})'$. In fact, the utilization of regularizer only changes the diagonal entries of $\hat{\boldsymbol{R}}_{k,n}$, which is $\hat{r}_0$. Asymptotically, $\hat{\boldsymbol{R}}_{k,n}^{\varepsilon}$ is equivalent to $\hat{\boldsymbol{R}}_{k,n}$ and $\hat{\boldsymbol{r}}_{k,0}^{\varepsilon}$ is equivalent to $\hat{\boldsymbol{r}}_{k,0}$ by the following argument: the regularizer vanishes as $n \to \infty$, i.e.,

$$\hat{r}_0^{\varepsilon} = \frac{1}{n}\sum_{t=1}^{n} Y_t^2 + \frac{\varepsilon e^{\mu(n)}}{n} \to \frac{1}{n}\sum_{t=1}^{n} Y_t^2 = \hat{r}_0 \quad a.s., \tag{2.47}$$

and therefore,

$$\hat{\boldsymbol{r}}_{k,0}^{\varepsilon} \to \hat{\boldsymbol{r}}_{k,0} \quad a.s., \quad \text{as } n \to \infty. \tag{2.48}$$

Lemma A.1 *Suppose that $E(\epsilon_t) = \kappa\sigma^4 < \infty$. If $n \to \infty$ and $k \to \infty$ such that $k^2/n \to 0$, then $\sqrt{n}(\hat{\boldsymbol{r}}_{k,0}^{\varepsilon} - \boldsymbol{r}_{k,0})$ converge to a Gaussian process with covariance function $\boldsymbol{\Sigma}^{\boldsymbol{r}}(i,j) = \lim_{k\to\infty} \Xi_k(i,j)$, where $\Xi_k(i,j)$ is the entry (i,j) of the $k \times k$ matrix $\boldsymbol{\Sigma}_k$.*

Proof of Lemma A.1 By (2.47) and (2.48), $\hat{\boldsymbol{r}}_{k,0}^{\varepsilon}$ is equivalent to $\hat{\boldsymbol{r}}_{k,0}$ as $n \to \infty$. Note that the assumption applied here on $\{\epsilon_t\}$, which assumes $\{\epsilon_t\} \sim i.i.d\ (0, \sigma^2)$ and $E(\epsilon_t^4) = \kappa\sigma^4 < \infty$, is a special case of the assumption applied in Lau et al. (2002). Hence, by Theorem 1 of Lau et

al. (2002), $\sqrt{n}(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0} - \boldsymbol{r}_{k,0})$ is asymptotically normally distributed with mean zero and covariance matrix $\boldsymbol{\Sigma}_k$, where $\boldsymbol{\Sigma}_k = \{\sigma^{\varepsilon}_{ij}\}^{k}_{i,j=0}$ and

$$\sigma^{\varepsilon}_{ij} = \lim_{n\to\infty} E\{n(\hat{r}^{\varepsilon}_i - r_i)(\hat{r}^{\varepsilon}_j - r_j)\}. \tag{2.49}$$

For any $j = 0, \ldots, k$, the estimation error of regularized sample ACVF estimate is given by

$$\begin{aligned}
\hat{r}^{\varepsilon}_j - r_j \quad &= \frac{j}{n}\sum_{s=1}^{q}\frac{\rho_s^2}{2}\cos(j\omega_s) + \frac{1}{n}\sum_{s=1}^{q}\rho_s^2\cos\big((n-1)\omega_s + 2\phi_s\big)\frac{\sin\big((n-j)\omega_s\big)}{2\sin\omega_s} \\
&\quad + \frac{1}{n}\sum_{t=1}^{n-j}X_{t+j}\epsilon_t + \frac{1}{n}\sum_{t=1}^{n-j}X_t\epsilon_{t+j} + \frac{1}{n}\sum_{t=1}^{n-j}\epsilon_t\epsilon_{t+j} + \frac{\delta_{j,0}\varepsilon e^{\mu(n)}}{n} - \delta_{j,0}\sigma^2 \\
&= A_{1j} + A_{2j} + A_{3j} + A_{4j} + A_{5j} + \frac{\delta_{j,0}\varepsilon e^{\mu(n)}}{n} - \delta_{j,0}\sigma^2.
\end{aligned} \tag{2.50}$$

Notice that as $n \to \infty$, $k \to \infty$ such that $k^2/n \to 0$, $nA_{1i}A_{1j} \to 0$ for $i, j = 0, \ldots, k$, Therefore, when $i, j \neq 0$, if $k^2/n \to 0$, we obtain

$$\begin{aligned}
\sigma^{\varepsilon}_{ij} &= \lim_{n\to\infty}\Big\{E\big(nA_{3i}A_{3j}\big) + E\big(nA_{4i}A_{4j}\big) + E\big(nA_{3i}A_{4j}\big) + E\big(nA_{4i}A_{3j}\big) + E\big(nA_{5i}A_{5j}\big)\Big\} \\
&= \delta_{i,j}\sigma^4 + \sigma^2\sum_{s=1}^{q}2\rho_s^2\cos(\omega_s i)\cos(\omega_s j).
\end{aligned}$$

When $i = j = 0$,

$$\sigma^{\varepsilon}_{00} = (\kappa - 1)\sigma^4 + \sigma^2\sum_{s=1}^{q}2\rho_s^2, \tag{2.51}$$

and the result follows.

Proof of Theorem 2.3.3 Since $\sqrt{n}(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0} - \boldsymbol{r}_{k,0})$ converges in distribution as stated by Lemma A.1, it follows the result of Serfling (1980) that $\hat{\boldsymbol{r}}^{\varepsilon}_{k,0} = \boldsymbol{r}_{k,0} + O(1/\sqrt{n})$. Define the following quantities:

- $g(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}) = (\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}\hat{\boldsymbol{r}}_k = \hat{\boldsymbol{\tau}}_{k,n}$ and $g(\boldsymbol{r}_{k,0}) = (\boldsymbol{R}_k)^{-1}\boldsymbol{r}_k = \boldsymbol{\tau}_k$,

- $\boldsymbol{\Delta}_{k,i}$= $(k \times k)$-matrix with $\pm i^{th}$ off-diagonal elements equal to 1, and 0 otherwise,
- $\boldsymbol{\vartheta}_{k,i}$= $(k \times 1)$-vector with $\pm i^{th}$ element equal to 1, and 0 otherwise.

Note in particular that $\boldsymbol{\Delta}_{k,0}$ is the identity matrix and $\boldsymbol{\Delta}_{k,k}$ is a zero matrix. In the view of matrix derivative (see, e.g., Gradshteyn et al., 2000; Lau et al., 2002),

$$\frac{\partial(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}}{\partial \hat{r}^{\varepsilon}_{i}} = -(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}\boldsymbol{\Delta}_{k,i}(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1} \text{ and } \frac{\partial \hat{\boldsymbol{r}}_k}{\partial \hat{r}^{\varepsilon}_{i}} = \boldsymbol{\vartheta}_{k,i}, \quad i = 0, 1, \ldots, k. \tag{2.52}$$

Thus, by the chain rule,

$$\begin{aligned}
\frac{\partial g(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0})}{\partial \hat{r}^{\varepsilon}_{i}}\bigg|_{\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}=\boldsymbol{r}_{k,0}} &= \left\{ \left(\frac{\partial(\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1}}{\partial \hat{r}^{\varepsilon}_{i}} \right) \hat{\boldsymbol{r}}_k + (\hat{\boldsymbol{R}}^{\varepsilon}_{k,n})^{-1} \left(\frac{\partial \hat{\boldsymbol{r}}_k}{\partial \hat{r}^{\varepsilon}_{i}} \right) \right\} \bigg|_{\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}=\boldsymbol{r}_{k,0}} \\
&= -(\boldsymbol{R}_k)^{-1}\boldsymbol{\Delta}_{k,i}(\boldsymbol{R}_k)^{-1}\boldsymbol{r}_k + (\boldsymbol{R}_k)^{-1}\boldsymbol{\vartheta}_{k,i} \\
&= -(\boldsymbol{R}_k)^{-1}(\boldsymbol{\Delta}_{k,i}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,i}).
\end{aligned} \tag{2.53}$$

Applying the Taylor expansion,

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) = \sqrt{n}\big\{ g(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}) - g(\boldsymbol{r}_{k,0}) \big\} &= \sqrt{n} \sum_{i=0}^{k} \frac{\partial g(\hat{\boldsymbol{r}}^{\varepsilon}_{k,0})}{\partial \hat{r}^{\varepsilon}_{i}}\bigg|_{\hat{\boldsymbol{r}}^{\varepsilon}_{k,0}=\boldsymbol{r}_{k,0}} (\hat{r}^{\varepsilon}_{i} - r_i) + o(1) \\
&= -\sum_{i=0}^{k} (\boldsymbol{R}_k)^{-1}(\boldsymbol{\Delta}_{k,i}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,i})\sqrt{n}(\hat{r}^{\varepsilon}_{i} - r_i) + o(1) \\
&= -(\boldsymbol{R}_k)^{-1}[\boldsymbol{\tau}_k, (\boldsymbol{\Delta}_{k,1}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,1}), \ldots, (\boldsymbol{\Delta}_{k,k}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,k})]\sqrt{n}(\hat{r}^{\varepsilon}_{i} - r_i) + o(1).
\end{aligned} \tag{2.54}$$

Let $a_i = 0$ for $i < 0$ and $i > k$. Note that

$$\boldsymbol{\Delta}_{k,i}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,i} = (a_{1+i}, a_{2+i}, \ldots, a_{k+i})' - (a_{1-i}, a_{2-i}, \ldots, a_{k-i})'. \tag{2.55}$$

Therefore, $[\boldsymbol{\tau}_k, (\boldsymbol{\Delta}_{k,1}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,1}), \ldots, (\boldsymbol{\Delta}_{k,k}\boldsymbol{\tau}_k - \boldsymbol{\vartheta}_{k,k})] = \boldsymbol{M}_k$ and the result follows by Lemma A.1. □

Proof of Theorem 2.3.4 Let $\{\hat{\beta}_s e^{\pm i\hat{\omega}_{k,s}}\}_{s=1}^{q}$ denote the $2q$ roots of $\hat{a}(z)$ which are closest to the unit circle. Applying the same arguments as in Stoica et al. (1989) and taking into account the

results on asymptotic consistency and normality of $\hat{\boldsymbol{\tau}}_{k,n}$ and Theorem 2, we obtain that $\hat{\omega}_{k,s}$ is close to ω_s, $s = 1, \ldots, q$, and $\hat{\beta}_s$ is close to $\beta_s = 1$ for sufficiently large n. Hence, the following Taylor expansion holds under regularity conditions:

$$\begin{aligned} 0 = \text{Re}\{\hat{a}(\hat{\beta}_s e^{i\hat{\omega}_{k,s}})\} &= \text{Re}\{\hat{a}(e^{i\omega_s})\} + \left.\frac{\partial \text{Re}\{\hat{a}(\beta e^{i\omega})\}}{\partial \beta}\right|_{\beta=1,\omega=\omega_s} (\hat{\beta}_s - \beta_s) \\ &+ \left.\frac{\partial \text{Re}\{\hat{a}(\beta e^{i\omega})\}}{\partial \omega}\right|_{\beta=1,\omega=\omega_s} (\hat{\omega}_{k,s} - \omega_s) + O(1/n), \end{aligned} \tag{2.56}$$

$$\begin{aligned} 0 = \text{Im}\{\hat{a}(\hat{\beta}_s e^{i\hat{\omega}_{k,s}})\} &= \text{Im}\{\hat{a}(e^{i\omega_s})\} + \left.\frac{\partial \text{Im}\{\hat{a}(\beta e^{i\omega})\}}{\partial \beta}\right|_{\beta=1,\omega=\omega_s} (\hat{\beta}_s - \beta_s) \\ &+ \left.\frac{\partial \text{Im}\{\hat{a}(\beta e^{i\omega})\}}{\partial \omega}\right|_{\beta=1,\omega=\omega_s} (\hat{\omega}_{k,s} - \omega_s) + O(1/n), \end{aligned} \tag{2.57}$$

where

$$\begin{aligned} \left.\frac{\partial \text{Re}\{\hat{a}(\beta e^{i\omega})\}}{\partial \beta}\right|_{\beta=1,\omega=\omega_s} &= (\cos\omega_s, 2\cos 2\omega_s, \ldots, k\cos k\omega_s)\hat{\boldsymbol{\tau}}_{k,n}, \\ \left.\frac{\partial \text{Re}\{\hat{a}(\beta e^{i\omega})\}}{\partial \omega}\right|_{\beta=1,\omega=\omega_s} &= -(\sin\omega_s, 2\sin 2\omega_s, \ldots, k\sin k\omega_s)\hat{\boldsymbol{\tau}}_{k,n}, \\ \left.\frac{\partial \text{Im}\{\hat{a}(\beta e^{i\omega})\}}{\partial \beta}\right|_{\beta=1,\omega=\omega_s} &= (\sin\omega_s, 2\sin 2\omega_s, \ldots, k\sin k\omega_s)\hat{\boldsymbol{\tau}}_{k,n}, \\ \left.\frac{\partial \text{Im}\{\hat{a}(\beta e^{i\omega})\}}{\partial \omega}\right|_{\beta=1,\omega=\omega_s} &= (\cos\omega_s, 2\cos 2\omega_s, \ldots, k\cos k\omega_s)\hat{\boldsymbol{\tau}}_{k,n}. \end{aligned} \tag{2.58}$$

By Theorem 2.3.3, as $k \to \infty$ and $T \to \infty$ such that $k^2/n \to 0$, $\sqrt{n}(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k)$ converges in distribution and thus it follows the result of Serfling (1980) that $(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) = O(1/\sqrt{n})$. Also, by Theorem 1 of Stoica et al. (1987), $(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) = O(1/k^2)$. Hence, we obtain

$$\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k^* = (\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) + (\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) = O(1/k^2) + O(1/\sqrt{n}). \tag{2.59}$$

Since $k^2/n \to 0$, the dominant term in (2.56) is not affected if we replace $\hat{\boldsymbol{\tau}}_{k,n}$ by $\boldsymbol{\tau}_k^*$, which is

$$\begin{aligned} 0 &= \text{Re}\{\hat{a}(e^{i\omega_s})\} + \theta_s(\hat{\beta}_s - \beta_s) - \psi_s(\hat{\omega}_{k,s} - \omega_s) + O(1/n), \\ 0 &= \text{Im}\{\hat{a}(e^{i\omega_s})\} + \psi_s(\hat{\beta}_s - \beta_s) + \theta_s(\hat{\omega}_{k,s} - \omega_s) + O(1/n). \end{aligned} \tag{2.60}$$

Since $a^*(e^{i\omega_s}) = 0$,

$$\begin{aligned} \text{Re}\{\hat{a}(e^{i\omega_s})\} &= \text{Re}\{\hat{a}(e^{i\omega_s}) - a^*(e^{i\omega_s})\} = \boldsymbol{h}_s'(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k^*), \\ \text{Im}\{\hat{a}(e^{i\omega_s})\} &= \text{Im}\{\hat{a}(e^{i\omega_s}) - a^*(e^{i\omega_s})\} = \boldsymbol{g}_s'(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k^*). \end{aligned} \tag{2.61}$$

Substituting (2.61) into (2.60), we obtain

$$(\hat{\omega}_{k,s} - \omega_s) = \frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k^*) + O(1/n). \tag{2.62}$$

Equivalently,

$$(\hat{\omega}_{k,s} - \omega_s) = \frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) + \frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) + O(1/n). \tag{2.63}$$

By the result of Stoica et al. (1987) Theorem 1,

$$(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) = O(1/k^2), \quad \theta_s/k = -1/2 + O(1/k), \quad \text{and} \quad \psi_s/k = O(1/k). \tag{2.64}$$

Substituting (2.64) into $(\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s')(\theta_s^2 + \psi_s^2)^{-1}(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*)$, we obtain

$$\frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\boldsymbol{\tau}_k - \boldsymbol{\tau}_k^*) = O(1/k^3). \tag{2.65}$$

Therefore,

$$(\hat{\omega}_{k,s} - \omega_s) = \frac{\psi_s \boldsymbol{h}_s' - \theta_s \boldsymbol{g}_s'}{\theta_s^2 + \psi_s^2}(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) + O(1/k^3) + O(1/n), \tag{2.66}$$

or equivalently,

$$\sqrt{n}(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}) = \sqrt{n}\boldsymbol{F}\boldsymbol{G}(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k) + O(\sqrt{n}/k^3) + O(1/\sqrt{n}). \tag{2.67}$$

If $k^2 \geqslant cn^{1-\delta}$, for $0 < \delta < 2/3$, then $\sqrt{n}/k^3 \to 0$, and $O(\sqrt{n}/k^3) \to 0$. Also, as $n \to \infty$, $O(1/\sqrt{n}) \to 0$, so we have

$$\sqrt{n}(\hat{\boldsymbol{\omega}}_k - \boldsymbol{\omega}) = \sqrt{n}\boldsymbol{F}\boldsymbol{G}(\hat{\boldsymbol{\tau}}_{k,n} - \boldsymbol{\tau}_k). \tag{2.68}$$

By Theorem 2.3.3, the result follows if $k^{3/2} \geqslant cn^{1-\delta}$, for $0 < \delta < 2/3$ such that $k^2/n \to 0$. □

In the rest of the discussion, we justify condition (2) in Conjecture 1.

Theorem A.1 *For any $\delta \in (0,1)$ and $\alpha > 0$, there exists a positive constant ζ_2 such that*

$$P\left\{\frac{1}{n}\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} \geq \zeta_2 \boldsymbol{I}_k\right\} \geq 1 - \frac{\alpha}{n^{\delta}}.$$

The main idea of proving Theorem A.1 is to divide $n^{-1}\hat{\boldsymbol{R}}^{\varepsilon}_{k,n}$ into three blocks as

$$\frac{1}{n}\hat{\boldsymbol{R}}^{\varepsilon}_{k,n} = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\Phi}_{k,t}\boldsymbol{\Phi}'_{k,t} + \frac{\varepsilon}{n}\boldsymbol{\Lambda_k} = Q_{1,n} + Q_{2,n} + Q_{3,n} \tag{2.69}$$

such that each block is dominated either by a non-zero observation data set or by the coefficients of the regularizer $\boldsymbol{\Lambda}_{k,n}$. Assume that $N = N(n)$ is a deterministic function of n such that $N < n$. The vector $\boldsymbol{\Phi}_{k,t}$ can be expressed in a state-space form as

$$\boldsymbol{\Phi}_{k,t} = \boldsymbol{A}^N\boldsymbol{\Phi}_{k,t-N} + \sum_{j=0}^{N-1}\boldsymbol{A}^j\boldsymbol{B}\nu_{k,t-j}, \tag{2.70}$$

where $\boldsymbol{A}$ is a $k \times k$ matrix and $\boldsymbol{B}$ is a $k \times 1$ vector of the forms

$$\boldsymbol{A} = \begin{pmatrix} -a_1 & -a_2 & \dots & -a_{k-1} & -a_k \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Let $b_1, \dots, b_p$ denote all the distinct eigenvalues of $\boldsymbol{A}$ and m_i is the multiplicity of b_i ($\sum_{i=1}^{p} m_i = k$). Let

$$\bar{b} = \max_{1\leq i\leq p}|b_i|, \quad \varpi = \max\{m_i : |b_i| = \bar{b}\}. \tag{2.71}$$

Note that $\lim_{k\to\infty}\bar{b}\to 1$. We show that $Q_{1,n}$, $Q_{2,n}$ and $Q_{3,n}$ are bounded from below as follows:

$$\begin{aligned}
Q_{1,n} &= \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{A}^{N}\boldsymbol{\Phi}_{k,t-N}\boldsymbol{\Phi}_{k,t-N}'\boldsymbol{A}^{*N}+\frac{\varepsilon_1}{n}\boldsymbol{\Lambda_k}\geq\frac{\varepsilon_1}{n}\boldsymbol{\Lambda_k},\\
Q_{2,n} &= 2\sum_{j=0}^{N-1}\boldsymbol{A}^{N}\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j}\right)\boldsymbol{B}'\boldsymbol{A}^{*j}+\frac{\varepsilon_2}{n}\boldsymbol{\Lambda_k}\geq -q_{2,n}\boldsymbol{I}_k,\\
Q_{3,n} &= \sum_{j=0}^{N-1}\sum_{l=0}^{N-1}\boldsymbol{A}^{j}\boldsymbol{B}\left(\frac{1}{n}\sum_{t=1}^{n}\nu_{k,t-j}\nu_{k,t-l}\right)\boldsymbol{B}'\boldsymbol{A}^{*l}\geq\sigma^2 U_N-q_{3,n}\boldsymbol{I}_k,
\end{aligned}$$

where $U_N=\sum_{j=0}^{N-1}\boldsymbol{A}^{j}\boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}^{*j}$; the constants $\varepsilon_1,\varepsilon_2>0$ are chosen such that $\varepsilon_1+\varepsilon_2=\varepsilon$; $\boldsymbol{A}^*$ denote the complex conjugate transpose of $\boldsymbol{A}$, i.e., $\boldsymbol{A}^*=\bar{\boldsymbol{A}}'$.

The next lemma investigates the limiting behavior of $n^{-1}\sum_{t=1}^{n}Y_t\nu_{k,t+h}$ and $n^{-1}\sum_{t=1}^{n}\nu_{k,t}\nu_{k,t+h}$. The results are useful for identifying the lower bounds of $Q_{2,n}$ and $Q_{3,n}$. First of all, let us introduce the following notations: for $0\leq h\leq n-1$,

$$\gamma_{k,h}=E(Y_t\nu_{k,t+h}),\quad \lambda_{k,h}=E(\nu_{k,t}\nu_{k,t+h})\quad\text{and}\quad \tilde{r}_h=n^{-1}\sum_{t=1}^{n}Y_tY_{t+h}.$$

Lemma A.2 *For $0\leq h\leq n-1$, as $k\to\infty$,*

(1) $\lim_{n\to\infty}nE\Big(n^{-1}\sum_{t=1}^{n}Y_t\nu_{k,t+h}-\gamma_{k,h}\Big)^2<\kappa_1$,

(2) $\lim_{n\to\infty}nE\Big(n^{-1}\sum_{t=1}^{n}\nu_{k,t}\nu_{k,t+h}-\lambda_{k,h}\Big)^2<\kappa_2$,

where κ_1 and κ_2 are positive constants.

Proof of Lemma A.2 Since $\nu_{k,t}=Y_t+\sum_{j=1}^{k}a_jY_{t-j}$, we can express $\frac{1}{n}\sum_{t=1}^{n}Y_t\nu_{k,t+h}$ and $\gamma_{k,h}$ respectively as

$$\frac{1}{n}\sum_{t=1}^{n}Y_t\nu_{k,t+h}=\frac{1}{n}\sum_{t=1}^{n}Y_tY_{t+h}+\sum_{j=1}^{k}a_j\left(\frac{1}{n}\sum_{t=1}^{n}Y_tY_{t+h-j}\right)=r_h+\sum_{j=1}^{k}a_j\tilde{r}_{h-j},\qquad(2.72)$$

and

$$\gamma_{k,h} = E\bigg(Y_t Y_{t+h} + \sum_{j=1}^{k} a_j Y_t Y_{t+h-j}\bigg) = r_h + \sum_{j=1}^{k} a_j r_{h-j}. \tag{2.73}$$

Thus we obtain, for $0 \le h \le n-1$,

$$\begin{aligned} & nE\bigg(\frac{1}{n}\sum_{t=1}^{n} Y_t \nu_{k,t+h} - \gamma_{k,h}\bigg)^2 = nE\bigg((\tilde{r}_h - r_h) + \sum_{j=1}^{k} a_j(\tilde{r}_{h-j} - r_{h-j})\bigg)^2 \\ = \;& nE(\tilde{r}_h - r_h)^2 + n\sum_{j=1}^{k} a_j^2 E(\tilde{r}_{h-j} - r_{h-j})^2 + 2n\sum_{i=1}^{k}\sum_{j=i+1}^{k} a_i a_j E\Big((\tilde{r}_{h-i} - r_{h-i})(\tilde{r}_{h-j} - r_{h-j})\Big) \\ + \;& 2n\sum_{j=1}^{k} a_j E\Big((\tilde{r}_h - r_h)(\tilde{r}_{h-j} - r_{h-j})\Big) = T_{1,h} + T_{2,h} + T_{3,h} + T_{4,h}. \end{aligned} \tag{2.74}$$

By Lemma 1 of Li et al.(1994) and equation (2.10), as $k \to \infty$, we have

$$\begin{aligned} \lim_{n\to\infty} T_{1,h} &= \sigma^4 + 2\sigma^2 \sum_{l=1}^{q} \rho_l^2 \cos^2(h\omega_l) \le \vartheta_1, \\ \lim_{n\to\infty} T_{2,h} &= \frac{4}{k^2}\sum_{j=1}^{k}\sum_{s=1}^{q} \cos^2(j\omega_s)\bigg(\sigma^4 + 2\sigma^2\sum_{l=1}^{q}\rho_l^2\cos^2\big((h-j)\omega_l\big)\bigg) \sim \frac{\vartheta_2}{k} = 0, \\ \lim_{n\to\infty} T_{3,h} &= \frac{16\sigma^2}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\bigg(\sum_{s=1}^{q}\sum_{s'=1}^{q}\sum_{l=1}^{q}\rho_l^2\cos(i\omega_s)\cos(j\omega_{s'})\cos\big((h-j-i)\omega_l\big)\cos(h\omega_l)\bigg) \le \vartheta_3, \\ \lim_{n\to\infty} T_{4,h} &= \frac{4}{k}\sum_{i=1}^{k}\sum_{s=1}^{q}\cos(j\omega_s)\bigg(\sigma^2\sum_{l=1}^{q}\rho_l^2\cos\big((h-j)\omega_l\big)\cos(h\omega_l)\bigg) \le \vartheta_4, \end{aligned}$$

where ϑ_1, ϑ_2, ϑ_3 and ϑ_4 are some constants. Therefore,

$$\lim_{n\to\infty} nE\bigg(\frac{1}{n}\sum_{t=1}^{n} Y_t\nu_{k,t+h} - \gamma_{k,h}\bigg)^2 \le \kappa_1, \tag{2.75}$$

where $\kappa_1 = \vartheta_1 + \vartheta_3 + \vartheta_4$.

Using similar arguments, we can show that, for $0 \leq h \leq n-1$,

$$\begin{aligned} & \lim_{n\to\infty} nE\left(\frac{1}{n}\sum_{t=1}^{n}\nu_{k,t}\nu_{k,t+h} - \lambda_{k,h}\right)^2 \\ = & \lim_{n\to\infty} nE\Bigg((\tilde{r}_h - r_h) + \sum_{j=1}^{k} a_j(\tilde{r}_{h-j} - r_{h-j}) + \sum_{i=1}^{k} a_i(\tilde{r}_{h-i} - r_{h-i}) \\ & \qquad + \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j(\tilde{r}_{h-i} - r_{h-i})(\tilde{r}_{h-j} - r_{h-j})\Bigg)^2 \leq \kappa_2, \end{aligned} \tag{2.76}$$

as $k \to \infty$, which completes the proof. □

The following two lemmas identify the lower bounds of $Q_{2,n}$ and $Q_{3,n}$ in probability.

Lemma A.3 *There exist* $\varrho_1,\ \varrho_2,\ \varrho_3 > 0$ *such that for any* $0 < \varphi < 1$, *if* $N(n) \leq \varrho_1(\varphi n^{1-\varphi})^{1/3}$, *then*

$$P\left\{q_{2,n} \leq \frac{\varrho_2 N^{2\varpi+1/2}}{n^{(1-\varphi)/2}}\right\} \geq 1 - \frac{\varrho_3}{n^{\varphi}}.$$

Proof of Lemma A.3 Note that $\{Y_t, t \in \mathbb{Z}\}$ is weakly stationary, so $E\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j} = E\boldsymbol{\Phi}_{k,j}\nu_{k,N}$, which can be expressed as

$$E\boldsymbol{\Phi}_{k,j}\nu_{k,N} = \left(\gamma_{k,N-j}, \gamma_{k,N-j+1}, \ldots, \gamma_{k,N-j+k-1}\right)'. \tag{2.77}$$

Next, let us re-write $Q_{2,n}$ as

$$\begin{aligned} Q_{2,n} &= \underbrace{2\sum_{j=0}^{N-1}\boldsymbol{A}^N\left(\frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j} - E\boldsymbol{\Phi}_{k,j}\nu_{k,N}\right)\right)\boldsymbol{B}'\boldsymbol{A}^{*j} + \frac{\varepsilon_2}{n}\boldsymbol{\Lambda_k}}_{:=H_{1,n}} \\ &- \underbrace{2\sum_{j=0}^{N-1}\boldsymbol{A}^N(E\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j})\boldsymbol{B}'\boldsymbol{A}^{*j}}_{:=H_{2,n}}. \end{aligned} \tag{2.78}$$

First, we investigate the asymptotic behavior of $H_{1,n}$. Let $q(n)$ be a positive function of sample size n, which will be specified later in this proof. Clearly, the diagonal operator $q\boldsymbol{I}_k + \varepsilon_2 n^{-1}\boldsymbol{\Lambda_k}$ is positive definite. Denote its square root as $S_q = \text{diag}\{s_i\}_{i=1}^k$ where $s_i = (q + \varepsilon_2 n^{-1}e^{\mu i})^{1/2}$. In order to show $H_{1,n} + q\boldsymbol{I}_k \geq 0$, it is sufficient to verify $\|S_q^{-1}(H_{1,n} + q\boldsymbol{I}_k)S_q^{-1} - \boldsymbol{I}_k\| \leq 1$, Note that

$$\begin{aligned} & \left\|S_q^{-1}(H_{1,n} + q\boldsymbol{I}_k)S_q^{-1} - \boldsymbol{I}_k\right\| \\ = & \sup_{\|c\|=1} \left\| S_q^{-1} 2\boldsymbol{A}^N \sum_{j=0}^{N-1} \left(\frac{1}{n}\sum_{t=1}^{n} \left(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j} - E\boldsymbol{\Phi}_{k,j}\nu_{k,N} \right) \right) \boldsymbol{B}'\boldsymbol{A}^{*j} S_q^{-1} c \right\| \leq 2\sqrt{p_1 p_2}, \end{aligned}$$

where

$$\begin{aligned} p_1 &= \sum_{j=0}^{N-1} \left\| S_q^{-1}\boldsymbol{A}^N \left(\frac{1}{n}\sum_{t=1}^{n} \left(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j} - E\boldsymbol{\Phi}_{k,j}\nu_{k,N} \right) \right) \right\|^2, \\ p_2 &= \sup_{\|c\|=1} \sum_{j=0}^{N-1} \left| c' S_q^{-1}\boldsymbol{A}^j\boldsymbol{B} \right|^2. \end{aligned}$$

Since $S_q \leq q^{1/2}\boldsymbol{I}_k$, the upper bound of p_2 is obtained by

$$p_2 \leq q^{-1} \sup_{\|c\|=1} \sum_{j=0}^{N-1} \left| c'\boldsymbol{A}^j\boldsymbol{B} \right|^2 \leq q^{-1}C \sum_{j=0}^{N-1} \|\boldsymbol{A}^j\|^2 \|\boldsymbol{B}\|^2 \leq q^{-1}C_1 \sum_{j=0}^{N-1} j^{2\varpi-2} \leq q^{-1}C_1 N^{2\varpi-1}, \tag{2.79}$$

where C and C_1 are positive constants.

Next, we derive an upper bound for p_1. By definition, $p_1 = \sum_{j=0}^{N-1} \|z_j\|^2$ where

$$z_j = \frac{1}{n} S_q^{-1}\boldsymbol{A}^N \sum_{t=1}^{n-N} \left(\boldsymbol{\Phi}_{k,t}\nu_{k,t+N-j} - E\boldsymbol{\Phi}_{k,j}\nu_{k,N} \right),$$

for $j = 0, \ldots, N-1$. Define the stochastic variables

$$\eta_{t,m} = \sum_{i=1}^{t} Y_i \nu_{k,i+m}, \;\; t \geq 1, \;\; m \geq 1. \tag{2.80}$$

Let operators P_N^-, P_N^+ split an arbitrary sequence $l = (l_0, \ldots, l_{k-1}) \in \mathbb{R}^k$ into two parts:

$$P_N^- l = (l_0, \ldots, l_{N-1}) \text{ and } P_N^+ = (l_N, l_{N+1}, \ldots, l_{k-1}).$$

Note that by definition $P_N^+ \boldsymbol{A}^N = (\boldsymbol{I}_{k-N}, \mathbf{0}_{(k-N)\times N})$, which implies

$$\begin{aligned} \left\| P_N^+ z_j \right\|^2 &= \frac{1}{n^2} \sum_{m=0}^{k-N-1} \left| \frac{1}{s_{m+N}} \sum_{t=1}^{n-N} \left(Y_{t-m} \nu_{k,t+N-j} - \gamma_{k,N-j+m} \right) \right|^2 \\ &= \frac{1}{n^2} \sum_{m=0}^{k-N-1} \left| \frac{1}{s_{m+N}} \left(\eta_{n-N-m,N-j+m} - E\eta_{n-N-m,N-j+m} \right) \right|^2. \end{aligned} \tag{2.81}$$

By applying Chebyshev's inequality, for $\varphi > 0$, we obtain

$$P\left\{ \frac{1}{t^{(1+\varphi)/2}} \left| \eta_{t,m} - E\eta_{t,m} \right| \leq \sqrt{\frac{C_2}{\varphi \alpha_m}} \right\} \geq 1 - \frac{\alpha_m \varphi}{C_2 t^{\varphi}} t E(t^{-1}\eta_{t,m} - \gamma_{k,m})^2. \tag{2.82}$$

By Lemma A.2 and choosing $\alpha_m = \alpha_0 m^{-2}$, the inequality (2.82) becomes

$$P\left\{ \left| \eta_{t,m} - E\eta_{t,m} \right| \leq \sqrt{\frac{C_2 m^2 t^{1+\varphi}}{\varphi \alpha_0}} \right\} \geq 1 - \frac{C_2'}{m^2 t^{\varphi}}, \tag{2.83}$$

where C_2 and C_2' are positive constants. Now let us substitute the upper bound of $|\eta_{t,m} - E\eta_{t,m}|$ obtained by the inequality (2.83) into (2.81),

$$\begin{aligned} \left\| P_N^+ z_j \right\|^2 &\leq \frac{C_2}{\varphi \alpha_0 n^2} \sum_{m=N}^{k-1} \frac{1}{s_{m+N}^2} (n - N - m)^{1+\varphi} (N - j + m)^2 \\ &\leq \frac{C_2}{\varphi \alpha_0 n^{1-\varphi}} \sum_{m=N}^{k-1} \frac{(N+m)^2}{q + \varepsilon_1 T^{-1} e^{\mu} m + n} \leq \frac{C_2}{\varphi \alpha_0 n^{1-\varphi}} \int_N^{\infty} \frac{x^2}{q + \varepsilon_1 T^{-1} e^{\mu x}} dx \\ &\leq \frac{C_2}{\varphi \alpha_0 n^{-\varphi}} \int_N^{\infty} x^2 e^{\mu x} dx \leq \frac{C_3 N^2}{\varphi \alpha_0 n^{1-\varphi}} \end{aligned} \tag{2.84}$$

for C_2 and C_3 positive constants, since $e^{\mu N} \geq n$. Next, we estimate $P_N^- z_j$ using a similar

approach.

$$\begin{aligned}
\left\|P_N^- z_j\right\|^2 &\leq \left\|S_q^{-1}\right\|^2 \left\|A^N\right\|^2 \frac{1}{n^2} \sum_{m=0}^{N-1} \left| \sum_{t=1}^{n-N} Y_{t-m}\nu_{t+N-j} - \gamma_{k,N-j+m} \right|^2 \\
&\leq \frac{C_4 N^{2\varpi-2}}{qn^2} \sum_{m=0}^{N-1} \left| \eta_{n-N-m,N-j+m} - E\eta_{n-N-m,N-j+m} \right|^2 \\
&\leq \frac{C_5 N^{2\varpi-2}}{q\varphi\alpha_0 n^2} \sum_{m=0}^{N-1} (n-N-m)^{1+\varphi}(N-j+m)^2 \\
&\leq \frac{C_5 N^{2\varpi-2}}{q\varphi\alpha_0 n^{1-\varphi}} \sum_{m=0}^{N-1} (N-j+m)^2 \leq \frac{C_5 N^{2\varpi+1}}{q\varphi\alpha_0 n^{1-\varphi}}, \qquad (2.85)
\end{aligned}$$

where C_4 and C_5 are positive constants. By definition $p_1 \leq N \max_{0\leq j\leq N-1} \Big(\|P_N^- z_j\|^2 + \|P_N^- z_j\|^2 \Big)$. Hence, for any $\alpha_0 > 0$,

$$P\left\{ p_1 \leq \frac{qC_3 N^3 + C_5 N^{2\varpi+2}}{q\varphi\alpha_0 n^{1-\varphi}} \right\} \geq 1 - \frac{C_2'}{m^2 n^\varphi} \geq 1 - \frac{C_2'}{n^\varphi}. \qquad (2.86)$$

The condition $2\sqrt{p_1 p_2} \leq 1$ is sufficient for $\|S_q^{-1}(H_{1,n} + q\boldsymbol{I}_k)S_q^{-1} - \boldsymbol{I}_k\| \leq 1$ and hence for $H_{1,n} + q\boldsymbol{I}_k \geq 0$. Thus, by (2.79) and (2.86), we require

$$\frac{4N^{4\varpi+1}}{q^2\varphi\alpha_0 n^{1-\varphi}} \left(C' + \frac{C''q}{N^{2\varpi-1}} \right) \leq 1, \qquad (2.87)$$

for $C' = C_1 C_3$ and $C'' = C_1 C_5$, which holds under two conditions:

$$q = q(n) = \frac{4\sqrt{C''} N^{2\varpi+1/2}}{n^{(1-\varphi)/2}\sqrt{2\varphi\alpha_0}} \text{ and } C'' N^{2\varpi-1} \geq C'q. \qquad (2.88)$$

The first equation is the definition of $q(n)$. The second equation holds if $N = N(n)$ satisfies

$$N \leq C_6 \Big(2\varphi\alpha_0 n^{1-\varphi} \Big)^{1/3}, \qquad (2.89)$$

where $C_6 = (\sqrt{C''}/4C')^{2/3}$. Therefore, for any $\alpha_0 > 0$ if $N \leq C_6 \Big(2\varphi\alpha_0 n^{1-\varphi} \Big)^{1/3}$, then

$$P\left\{ H_{1,n} + \frac{C_7 N^{2\varpi+1/2}}{n^{(1-\varphi)/2}} \boldsymbol{I}_k \geq H_{1,n} + q_{2,n}\boldsymbol{I}_k \geq 0 \right\} \geq 1 - \frac{C_2'}{n^\varphi}. \qquad (2.90)$$

where $C_7 = \sqrt{8C''/(\varphi\alpha_0)}$.

Second, we investigate the asymptotic behavior of $H_{2,n}$. By substituting (2.77) into $H_{2,n}$, we obtain

$$H_{2,n} = 2\sum_{j=0}^{N-1} \boldsymbol{A}^N \big(\gamma_{k,N-j}, \ldots, \gamma_{k,N-j+k-1}\big)' \boldsymbol{B}' \boldsymbol{A}^{*j}. \tag{2.91}$$

Since $\gamma_{k,h} = 0$ for $h = 1, \ldots, k$, when $j = 0$,

$$\boldsymbol{A}^N \big(\gamma_{k,N-j}, \gamma_{k,N-j+1}, \ldots, \gamma_{k,N-j+k-1}\big)' \boldsymbol{B}' \boldsymbol{A}^{*j} = \boldsymbol{A}^N \big(\underbrace{0, \ldots, 0}_{k-N+1 \text{ zeros}}, \gamma_{k,k+1}, \ldots, \gamma_{k,N+k-1}\big)' \boldsymbol{B}'$$

$$= \begin{pmatrix} \psi_{11}(\boldsymbol{\tau}_k) & \psi_{12}(\boldsymbol{\tau}_k) & \ldots & \psi_{1,k-N}(\boldsymbol{\tau}_k) & \ldots & \psi_{1k}(\boldsymbol{\tau}_k) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \psi_{N-1,1}(\boldsymbol{\tau}_k) & \psi_{N-1,2}(\boldsymbol{\tau}_k) & \ldots & \psi_{N-1,k-N}(\boldsymbol{\tau}_k) & \ldots & \psi_{N-1,k}(\boldsymbol{\tau}_k) \\ -a_1 & -a_2 & \ldots & -a_{k-N} & \ldots & -a_k \\ 1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & \ldots & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ \gamma_{k,k+1} \\ \vdots \\ \gamma_{k,N+k-1} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}'$$

$$= \begin{pmatrix} \sum_{m=1}^{N-1} \psi_{1,k-N+1+m}(\boldsymbol{\tau}_k)\gamma_{k,k+m} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{m=1}^{N-1} \psi_{N-1,k-N+1+m}(\boldsymbol{\tau}_k)\gamma_{k,k+m} & 0 & \ldots & 0 \\ -\sum_{m=1}^{N-1} a_{k-N+1+m}\gamma_{k,k+m} & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix},$$

where $\psi_{il}(\boldsymbol{\tau}_k)$ at $(i,j)^{th}$ entry of $\boldsymbol{A}^N$ is a polynomial of $\boldsymbol{\tau}_k = -(a_1, \ldots, a_k)$ of degree $N+1-i$

and hence by result (2.10), $\psi_{il}(\boldsymbol{\tau}_k) \leq \varsigma_1/k$, $1 \leq i, l \leq k$, for some constant ς_1. Similarly, $a_i \leq \varsigma_2/k$, $1 \leq i \leq k$, for some constant ς_2. Thus, we obtain

$$\left\| \boldsymbol{A}^N E(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t})\boldsymbol{B}' \right\| \leq \varrho_1 \sqrt{\frac{N(N-1)^2}{k^2}\gamma_{k,k+1}^2} \leq \varrho_2 N^{3/2}/k, \tag{2.92}$$

where ϱ_1 and ϱ_2 are positive constants. Since the number of non-zero terms in $E(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j})$ decreases as j increases,

$$\left\| \boldsymbol{A}^N E(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t})\boldsymbol{B}' \right\| = \max_{0\leq j\leq N-1} \left\| \boldsymbol{A}^N E(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j})\boldsymbol{B}'\boldsymbol{A}^{*j} \right\| \tag{2.93}$$

and therefore

$$\begin{aligned} \left\| H_{2,n} \right\| &\leq 2\sum_{j=0}^{N-1} \left\| \boldsymbol{A}^N E(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t-j})\boldsymbol{B}'\boldsymbol{A}^{*j} \right\| \leq 2(N-1) \left\| \boldsymbol{A}^N E(\boldsymbol{\Phi}_{k,t-N}\nu_{k,t})\boldsymbol{B}' \right\| \\ &\leq \varrho_3 N^{5/2}/k < \infty, \end{aligned} \tag{2.94}$$

for some positive constant ϱ_3.

Finally, by results (2.90) and (2.94), we obtain

$$P\left\{ Q_{2,n} + \frac{C_7 N^{2\varpi+1/2}}{n^{(1-\varphi)/2}}\boldsymbol{I}_k \geq Q_{2,n} + q_{2,n}\boldsymbol{I}_k \geq 0 \right\} \geq 1 - \frac{C_2'}{n^{\varphi}}, \tag{2.95}$$

which completes the proof. □

Lemma A.4

(1) There exist $\varrho_4, \varrho_5 > 0$ *such that for any* $0 < \beta < \alpha < 1$,

$$P\left\{ q_{3,n} \leq \frac{\varrho_5 N^{2\varpi-1}}{n^{\beta/2}} \right\} \geq 1 - \frac{\varrho_4}{n^{\alpha-\beta}}.$$

(2) There exist a constant $d > 0$ such that

$$P_N \sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}' \boldsymbol{A}^{*j} P_N \geq d P_N,$$

where P_m is a standard projector, $P_m c = (c_0, c_1, \ldots, c_{m-1}, 0, \ldots, 0)$ for $c = (c_0, c_1, \ldots, c_k)$.

Proof of Lemma A.4 Let us re-write $Q_{3,n}$ as

$$\begin{aligned} Q_{3,n} &= \sum_{j=0}^{N-1}\sum_{l=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}' \boldsymbol{A}^{*l} \left(\frac{1}{n} \sum_{t=1}^{n} \nu_{k,t-j}\nu_{k,t-l} - \lambda_{k,|l-j|} \right) + \sum_{j=0}^{N-1}\sum_{l=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}' \boldsymbol{A}^{*l} \lambda_{k,|l-j|} \\ &= G_{1,n} + G_{2,n}. \end{aligned} \tag{2.96}$$

By Chebyshev's inequality, for $0 < \beta < \alpha < 1$, we obtain

$$\begin{aligned} P\left\{ \frac{1}{n^{(1+\alpha)/2}} \left| \sum_{t=1}^{n} \nu_{k,t-j}\nu_{k,t-l} - \lambda_{k,|l-j|} \right| \leq \frac{1}{n^{\beta/2}} \right\} &\geq 1 - \frac{E\left(\frac{1}{n}\sum_{t=1}^{n} \nu_{k,t-j}\nu_{k,t-l} - \lambda_{k,|l-j|} \right)^2}{n^{\alpha-\beta}} \\ &\geq 1 - \frac{\varrho_4}{n^{\alpha-\beta}}. \end{aligned} \tag{2.97}$$

for some constant $\varrho_4 > 0$. The second inequality of (2.97) is due to Lemma A.2 and hence

$$\|G_{1,n}\| \leq \frac{C}{n^{\beta/2}} \sum_{j=0}^{N-1} \|\boldsymbol{A}^j\|^2 \|\boldsymbol{B}\|^2 \leq \frac{\varrho_5 N^{2\varpi-1}}{n^{\beta/2}}, \tag{2.98}$$

where C and $\varrho_5 > 0$ are positive constants. This implies that

$$P\left\{ G_{1,n} + \frac{\varrho_5 N^{2\varpi-1}}{n^{\beta/2}} \boldsymbol{I}_k \geq G_{1,n} + q_{3,n} \boldsymbol{I}_k \geq 0 \right\} \geq 1 - \frac{\varrho_4}{n^{\alpha-\beta}}, \tag{2.99}$$

which completes the proof of part (1).

Next, let us investigate the asymptotic behavior of $G_{2,n}$, which does not contain random variables. Using the result $r_{|l-j|} = \sum_{b=1}^{q} \frac{\rho_b^2}{2} \cos(|l-j|\omega_b) + \delta_{l,j}\sigma^2$ of Li et al. (1994) and the

complex exponential representation of cosine function, we obtain

$$\begin{aligned}
\lambda_{k,|l-j|} &= r_{|l-j|} + \sum_{s=1}^{k} a_s r_{|l-j-s|} + \sum_{m=1}^{k} a_m r_{|l-j+m|} + \sum_{s=1}^{k}\sum_{m=1}^{k} a_s a_m r_{|l-j-s+m|} \\
&= \mathrm{Re}\sum_{b=1}^{q} \frac{\rho_b^2}{2}\left(e^{i|l-j|\omega_b} + \sum_{s=1}^{k} a_s e^{i|l-j-s|\omega_b} + \sum_{m=1}^{k} a_m e^{i|l-j+m|\omega_b} + \sum_{s=1}^{k}\sum_{m=1}^{k} a_s a_m e^{i|l-j-s+m|\omega_b} \right) \\
&+ \sigma^2\left(\delta_{l,j} + \sum_{s=1}^{k} a_s \delta_{l-j,s} + \sum_{m=1}^{k} a_m \delta_{j-l,m} + \sum_{s=1}^{k}\sum_{m=1}^{k} a_s a_m \delta_{l-j,s-m} \right) \\
&= \mathrm{Re}\left(\sum_{b=1}^{q} \frac{\rho_b^2}{2} e^{i|l-j|\omega_b} \left| 1 + \sum_{s=1}^{k} a_s e^{is\omega_b} \right|^2 \right) + \sigma^2 \delta_{l,j} + O(1/k) \\
&= C_1 \sum_{b=1}^{q} \frac{\rho_b^2}{2} \cos(|l-j|\omega_b) + \sigma^2 \delta_{l,j} + O(1/k), \qquad (2.100)
\end{aligned}$$

for some constant $C_1 \geq 0$ and hence,

$$\begin{aligned}
G_{2,n} &= C_1 \sum_{b=1}^{q} \frac{\rho_b^2}{2} \sum_{j=0}^{N-1}\sum_{l=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}^{*l} \cos(|l-j|\omega_b) + \sigma^2 \sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}^{*j} + O(N^2/k) \\
&= C_1 \sum_{b=1}^{q} \frac{\rho_b^2}{2} \mathrm{Re}\left| \sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B} e^{ij\omega_b} \right|^2 + \sigma^2 \sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}^{*j} + O(N^2/k). \qquad (2.101)
\end{aligned}$$

Since $\sum_{b=1}^{q} \frac{\rho_b^2}{2} \mathrm{Re}\left| \sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B} e^{ij\omega_b} \right|^2 \geq 0$ and $\sigma^2 \sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}^{*j} > 0$, as $k \to \infty$ such that $N^2/k \to 0$,

$$G_{2,n} \geq \sigma^2 \sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}'\boldsymbol{A}^{*j} \geq 0. \qquad (2.102)$$

Notice that for $j = 1, \ldots, N-1$,

$$\boldsymbol{A}^j \boldsymbol{B} = \Big(\psi_1(\boldsymbol{\tau}_k), \ldots, \psi_j(\boldsymbol{\tau}_k), 1, \underbrace{0, \ldots, 0}_{k-j-1 \text{ zeros}} \Big)', \qquad (2.103)$$

where $\psi_i(\boldsymbol{\tau}_k)$ is a polynomial of $\boldsymbol{\tau}_k = -(a_1, \ldots, a_k)$ of degree $j - i + 1$. By the structure of $\boldsymbol{A}^j \boldsymbol{B}$, it is clear that $\left\{ \boldsymbol{B}, \boldsymbol{A}\boldsymbol{B}, \boldsymbol{A}^2\boldsymbol{B}, \ldots, \boldsymbol{A}^{N-1}\boldsymbol{B} \right\}$ are linearly independent and hence

$\sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}' \boldsymbol{A}^{*j}$ is positive definite. Therefore, there exist a constant d such that

$$P_N \sum_{j=0}^{N-1} \boldsymbol{A}^j \boldsymbol{B}\boldsymbol{B}' \boldsymbol{A}^{*j} P_N \geq d P_N,$$

which completes the proof of part (2).

Proof of Theorem A.1 Choose $\beta = 1-\varphi$ and $\alpha/n^{\delta} = \max\left\{\varrho_3/n^{\varphi}, \varrho_4/n^{\alpha-1+\varphi}\right\}$. By Lemma A.3 and A.4, we obtain

$$P\left\{\frac{1}{n}\hat{\boldsymbol{R}}_{k,n}^{\varepsilon} \geq \frac{\varepsilon_1}{n}\boldsymbol{\Lambda}_{\boldsymbol{k}} + d\sigma^2 P_N(n) - \frac{\left(\varrho_2 N(n)^{3/2}/\varrho_5 + 1\right)\varrho_5 N(n)^{2\varpi-1}}{n^{1-\varphi}}\right\} \geq 1 - \frac{\alpha}{n^{\delta}},$$

if $N(n) \leq \varrho_1(\varphi n^{1-\varphi})^{1/3}$. Let us fix $N(n) = \mu^{-1}\ln n$, then

$$\frac{\varepsilon_1}{n}\boldsymbol{\Lambda}_{\boldsymbol{k}} + d\sigma^2 P_N(n) \geq \zeta_2 \boldsymbol{I}_k, \quad \text{for } \zeta_2 = \min\left\{\varepsilon_1 e^{-\mu N}, d\sigma^2\right\}, \tag{2.104}$$

and the result of Theorem A.1 follows. □

Chapter 3

Sieve Bootstrap Prediction Intervals for Returns and Volatilities in GARCH

Measuring volatility plays an important role in assessing risk and uncertainty in financial markets. One of the core techniques for modeling volatility dynamics in empirical finance is the autoregressive conditional heteroscedastic (ARCH) model, introduced by Engle (1982). The pioneering idea of the ARCH approach is to view volatility as a linear function of previous squared returns. By adding a 'moving average' (MA) part, Bollerslev (1986) proposes incorporating available information on previous volatilities, which resulted in the Generalized ARCH (GARCH) model. There now exists a variety of modifications of the ARCH/GARCH approach: exponential GARCH (Nelson, 1991), nonlinear GARCH (Engle and Ng, 1993), integrated GARCH (Engle and Bollerslev, 1986), fractionally integrated GARCH (Baillie et al. 1996), long memory GARCH (Conrad and Karanasos, 2006), etc, and new extensions continue to appear frequently (see Bera and Higgins, 1993; Shephard, 1996; Tsay, 2002; Taylor, 2005; Bollerslev, 2008, and references therein for an overview). Although the sequence of volatili-

ties is typically unobservable, predicting volatility by ARCH/GARCH models is straightforward due to their functional structure. However, the existing literature mainly focuses on point forecasts of volatility (see Baillie and Bollerslev, 1992; Andersen and Bollerslev, 1998; Andersen et al., 2001; Engle and Patton, 2001; Poon, 2005, and references therein), and relatively little attention has been paid to constructing prediction intervals. Compared to point forecasts, prediction intervals provide extra assessment of the uncertainty associated with the corresponding point forecast, which can better guide risk management decisions. However, construction of prediction intervals requires knowledge of the distribution of the observed data, which is typically unknown in practice. Hence, data are usually assumed to follow some hypothetical distribution, and the resulting prediction interval can be adversely affected by departures from that assumption (Thombs and Schucany, 1990). An alternative is to employ distribution-free re-sampling techniques, e.g., the bootstrap. One of the most popular and efficient bootstrap procedures in a time series context is based on assessing the predictive error distribution by re-sampling residuals from the fitted model (Bühlmann, 2002; Politis, 2003; Härdle et al., 2003). In particular, Miguel and Olave (1999) propose constructing bootstrap-based prediction intervals of returns and volatilities by directly adding re-sampled residuals from an ARCH model to the respective point forecasts. Reeves (2005) suggests including an additional step of re-estimating the ARCH parameters for each bootstrapped realization of returns, which incorporates the uncertainty in sample parameter estimates. Pascual et al. (2006) combine and further extend these procedures by developing prediction intervals for both returns and volatilities from GARCH models, and the obtained prediction intervals are found to be well-calibrated, i.e., the number of observed data falling within a prediction interval coincides with the declared coverage. However, the discussed procedures are based on maximum likelihood (ML) estimation of ARCH/GARCH parameters and, hence, are computationally expensive.

In this chapter, we propose a novel, simple and efficient sieve-based bootstrap procedure to construct prediction intervals of returns and volatilities in ARCH/GARCH processes. The sieve bootstrap is a residual-based re-sampling technique widely applied in linear time series modeling due to its efficiency, low computational costs and non-restrictive assumptions (Kreiss, 1988; Bühlmann, 1997; Politis, 2003; Härdle et al., 2003; Pascual et al., 2004). The idea of the sieve bootstrap is to approximate an observed process by a linear model, typically autoregressive (AR), and to generate "new" realizations from the same model but with the re-sampled innovations. Notice that ARCH/GARCH equations can also be represented as AR/ARMA processes from the Box-Jenkins family of models. In particular, the squared return from an ARCH/GARCH model is a linear process that follows an AR/ARMA equation (Tsay, 2002; Box et al., 2008). Hence, we can also adopt a sieve bootstrap procedure for the ARCH/GARCH case, i.e. develop prediction intervals for squared returns and then apply probability transformations to construct the required prediction intervals for returns and volatility. Since the proposed approach involves only estimation of AR/ARMA models by least squares (LS), it significantly reduces the computational costs while providing sharp and well-calibrated prediction intervals for both returns and volatilities.

3.1 ARCH and GARCH models as AR and ARMA models

We start our discussion from a general class of GARCH(p, q) models and then consider ARCH(1) and GARCH $(1, 1)$ processes as examples.

3.1.1 Models and assumptions

Suppose $\{Y_t\}_{t\in\mathbb{Z}}$ follows a GARCH(p, q) process, $p, q \geq 1$,

$$Y_t = \sigma_t \epsilon_t \tag{3.1}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3.2}$$

where $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent, identically distributed (i.i.d.) random variables with $E(\epsilon_t) = 0$ and $E(\epsilon_t^2) = 1$; $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is a stochastic process assumed to be independent of $\{\sigma_t\}_{t\in\mathbb{Z}}$; α_0, α_i and β_j are unknown parameters satisfying $\alpha_0 \geq 0$, $\alpha_i \geq 0$ and $\beta_j \geq 0$, for $i = 1, \ldots, p$ and $j = 1, \ldots, q$. Let $m = \max(p, q)$. Throughout this paper, we assume that $\{Y_t\}_{t\in\mathbb{Z}}$ is weakly stationary, i.e., $\sum_{i=1}^{m}(\alpha_i + \beta_i) < 1$ is satisfied, where $\alpha_i = 0$ for $i > p$ and $\beta_i = 0$ for $i > q$ (Tsay, 2002). Further, we assume that the strict stationarity conditions of $\{Y_t\}_{t\in\mathbb{Z}}$ given in Bougerol and Picard (1992a,b) hold [1]. Note that in financial contexts, $\{Y_t\}_{t\in\mathbb{Z}}$ and $\{\sigma_t\}_{t\in\mathbb{Z}}$ are referred to as return and volatility processes respectively. If $q = 0$, $\{Y_t\}_{t\in\mathbb{Z}}$ is an ARCH(p) process. Let us

[1] Bougerol and Picard (1992a, 1992b) provide the necessary and sufficient conditions for the existence of a strictly stationary solution of (3.1) and (3.2). Let $\tau_n = (\beta_1 + \alpha_1\epsilon_n^2, \beta_2, \ldots, \beta_{q-1})' \in \mathbb{R}^{q-1}$, $\xi_n = (\epsilon_n^2, 0, \ldots, 0)' \in \mathbb{R}^{q-1}$, $\alpha = (\alpha_2, \ldots, \alpha_{p-1})' \in \mathbb{R}^{p-2}$. Let

$$A_n = \begin{bmatrix} \tau_n & \beta_q & \alpha & \alpha_p \\ I_{q-1} & 0 & 0 & 0 \\ \xi_n & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{bmatrix},$$

where I_k is a $k \times k$-identity matrix for $k \in \mathbb{Z}^+$. The top Lyapunov exponent γ_L associated with the sequence $\{A_n, -\infty < n < \infty\}$ is $\gamma_L = \inf_{0\leq n<\infty} \frac{1}{n+1} E \log \|A_0 A_1 \ldots A_n\|$, assuming that $E(\log \|A_0\|) < \infty$. (Here $\|M\| = \sup\{\|M\mathbf{x}\|_d / \|\mathbf{x}\|_d : \mathbf{x} \in \mathbb{R}^d, \mathbf{x} \neq 0\}$ and $\|\cdot\|_d$ is the Euclidean norm in $\mathbb{R}^d$.) Bougerol and Picard (1992a, 1992b) show that if $E(\log \|A_0\|) < \infty$ holds, then (3.1) and (3.2) have a unique stationary solution if and only if $\gamma_L < 0$.

now represent a GARCH (p, q) process in an autoregressive moving average (ARMA) form. If we denote $v_t = Y_t^2 - \sigma_t^2$, then

$$Y_t^2 = \alpha_0 + \sum_{i=1}^{m}(\alpha_i + \beta_i)Y_{t-i}^2 + v_t - \sum_{j=1}^{q}\beta_j v_{t-j}, \tag{3.3}$$

where $\{v_t\}_{t\in\mathbb{Z}}$ is white noise, but not i.i.d.in general. Under the strict stationary assumption of $\{Y_t\}_{t\in\mathbb{Z}}$, $\{v_t\}_{t\in\mathbb{Z}}$ is identically distributed. Note that if $q = 0$, formula (3.3) reduces to an AR(m) model. For example, let us consider the linear forms of two special cases of a GARCH(p, q) process. Suppose $\{Y_t\}_{t\in\mathbb{Z}}$ follows an ARCH (1):

$$Y_t = \sigma_t\epsilon_t \tag{3.4}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2. \tag{3.5}$$

Then, in view of (3.3), equations (3.4) and (3.5) can be expressed in the AR(1) form

$$Y_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + v_t. \tag{3.6}$$

Note that $\{Y_t\}_{t\in\mathbb{Z}}$ is both weakly and strictly stationary if $\alpha_1^2 \leq 1/3$ (Box et al., 2008; Tsay, 2002). Now suppose $\{Y_t\}_{t\in\mathbb{Z}}$ is a GARCH $(1, 1)$ process given by

$$Y_t = \sigma_t\epsilon_t \tag{3.7}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1\sigma_{t-1}^2. \tag{3.8}$$

Then (3.7) and (3.8) can be rewritten as the ARMA$(1, 1)$ form

$$Y_t^2 = \alpha_0 + (\alpha_1 + \beta_1)Y_{t-1}^2 + v_t - \beta_1 v_{t-1}, \tag{3.9}$$

assuming that $\alpha_1 + \beta_1 < 1$ to ensure the weak stationarity of $\{Y_t\}_{t\in\mathbb{Z}}$ (Box et al., 2008; Tsay, 2002). Nelson (1990) showed that $\{Y_t\}_{t\in\mathbb{Z}}$ is also strictly stationary if $E[\log(\beta_1 + \alpha_1\epsilon_t)] < 1$. The linear representation of the GARCH(p, q) process allows us to utilize the sieve bootstrap algorithm to construct prediction intervals for returns and volatilities.

3.1.2 Sieve bootstrap procedure of GARCH (p, q) process

The sieve bootstrap is proposed for estimating the distribution of a statistical quantity within a class of linear processes (Kreiss, 1988; Bühlmann, 1997). Given a sample of size n, the idea of the sieve bootstrap is to fit a sequence of AR models of order $p(n)$, where $p(n) \to \infty$ as $n \to \infty$, and then to construct a "new" bootstrap realization generated from the re-sampled residuals (Grenander, 1981). The asymptotic properties of the sieve bootstrap are studied by Bühlmann (1997), Bickel and Bühlmann (1999), Härdle et al., (2003), Politis, (2003) and Lahiri, (2003). Recently, the sieve bootstrap has been gaining popularity for constructing prediction intervals for linear processes. In particular, Thombs and Schucany (TS) (1990) and Cao et al. (1997) consider the performance of sieve bootstrap prediction intervals for finite AR(p) models, while Alonso et al. (2002, 2003) extend the sieve bootstrap algorithm to the AR(∞) model with absolutely summable coefficients and Pascual et al. (2004) apply the sieve bootstrap procedure to integrated ARMA (ARIMA) processes. Here we adopt the sieve bootstrap idea for developing prediction intervals of returns and volatility in GARCH(p, q) processes. Let $h = 1, \ldots, s$, $s \geq 1$, be a lead time. Denote by $\{Y_t\}_{t=1}^{n}$ a sample from the process $\{Y_t\}_{t\in\mathbb{Z}}$, where p and q are assumed to be known.[2] Then, we proceed with the following algorithm to construct prediction intervals for Y_{T+h} and σ^2_{T+h}:

- Step 1. Estimate the ARMA coefficients $\hat{\alpha}_0$, $\widehat{(\alpha_1 + \beta_1)}, \ldots, \widehat{(\alpha_m + \beta_m)}$, $\hat{\beta}_1, \ldots, \hat{\beta}_q$ from the representation (3.3) using the Least Square (LS) method. Then, calculate $\hat{\alpha}_i = \widehat{(\alpha_i + \beta_i)} - \hat{\beta}_i$, for $i = 1, \ldots, p$.

[2]Note that in practice, the order of the GARCH process is unknown. We can select the model order by the Akaike information criterion (AIC) (Akaike, 1974), the Bayesian information criterion (BIC) (Schwarz, 1978) or the corrected AIC (AICc) (Hurvich and Tsai, 1989).

- Step 2. Estimate the residuals $\{\hat{v}_t\}_{t=m+1}^{n}$ by

$$\hat{v}_t = Y_t^2 - \hat{\alpha}_0 - \sum_{i=1}^{m} \widehat{(\alpha_i + \beta_i)} Y_{t-i}^2 + \sum_{j=1}^{q} \hat{\beta}_j \hat{v}_{t-j}, \;\; t = m+1, \ldots, n. \tag{3.10}$$

  Set $\hat{v}_t = 0$, for $t = 1, \ldots, m$.

- Step 3. Center the estimated residuals using

$$\tilde{v}_t = \left(\hat{v}_t - \frac{1}{n-m} \sum_{t=m+1}^{n} \hat{v}_t \right), \;\; t = m+1, \ldots, n. \tag{3.11}$$

  The empirical distribution of the centered residuals $\{\tilde{v}_t\}_{t=m+1}^{n}$ is

$$\hat{F}_{v,n}(y) = \frac{1}{n-m} \sum_{t=m+1}^{n} \mathbf{1}_{\{\tilde{v}_t \leq y\}}. \tag{3.12}$$

- Step 4. Sample with replacement from $\hat{F}_{v,n}(y)$ to obtain the bootstrap error process $\{v_t^*\}_{t=1}^{n}$.

- Step 5. Construct a bootstrap sample of squared returns $\{Y_t^{2*}\}_{t=1}^{n}$ by

$$Y_t^{2*} = \hat{\alpha}_0 + \sum_{i=1}^{m} \widehat{(\alpha_i + \beta_i)} Y_{t-i}^{2*} + v_t^* - \sum_{j=1}^{q} \hat{\beta}_j v_{t-j}^*, \tag{3.13}$$

  where $Y_k^{2*} = \hat{\alpha}_0 / \{1 - \sum_{i=1}^{m} \widehat{(\alpha_i + \beta_i)}\}$ and $v_k^* = 0$ for $k \leq 0$. In practice we generate Y_t^{2*} of length $n + 150$ and then discard the first 150 generated values in order to minimize the effect of the initial values. The effect of initial values is negligible asymptotically (Kreiss and Franke, 1992).

- Step 6. Given $\{Y_t^{2*}\}_{t=1}^{n}$ from Step 5, estimate the coefficients $\hat{\alpha}_0^*$, $\widehat{(\alpha_1 + \beta_1)}^*, \ldots, \widehat{(\alpha_m + \beta_m)}^*$, $\hat{\beta}_1^*, \ldots, \hat{\beta}_q^*$ by LS and then calculate $\hat{\alpha}_i^* = \widehat{(\alpha_i + \beta_i)}^* - \hat{\beta}_i^*$, for $i = 1, \ldots p$. The bootstrap sample of volatility $\{\sigma_t^{2*}\}_{t=1}^{n}$ is obtained by

$$\sigma_t^{2*} = \hat{\alpha}_0^* + \sum_{i=1}^{p} \hat{\alpha}_i^* Y_{t-i}^{2*} + \sum_{j=1}^{q} \hat{\beta}_j^* \sigma_{t-j}^{2*}, \;\; t = m+1, \ldots, n \tag{3.14}$$

with $\sigma_t^{2*} = \hat{\alpha}_0/\{1 - \sum_{i=1}^{m}(\hat{\alpha}_i + \hat{\beta}_i)\}$, $t = 1, \ldots, m$.

- Step 7. Sample with replacement from $\hat{F}_{v,n}(y)$ to obtain the bootstrap prediction error process $\{v_{n+h}^*\}_{h=1}^{s}$, $s \geqslant 1$.

- Step 8. Let $Y_{n+h}^* = Y_{n+h}$, for $h \leqslant 0$. The h-step-ahead forecast of the squared return is given by

$$Y_{n+h}^{2*} = \hat{\alpha}_0^* + \sum_{i=1}^{m} \widehat{(\alpha_i + \beta_i)}^* Y_{n+h-i}^{2*} + v_{n+h}^* - \sum_{j=1}^{q} \hat{\beta}_j^* v_{n+h-j}^*, \tag{3.15}$$

  and the h-step-ahead forecast of volatility is obtained by

$$\sigma_{n+h}^{2*} = \hat{\alpha}_0^* + \sum_{i=1}^{p} \hat{\alpha}_i^* Y_{n+h-i}^{2*} + \sum_{j=1}^{q} \hat{\beta}_j^* \sigma_{n+h-j}^{2*}, \tag{3.16}$$

  for $h = 1, \ldots, s$.

- Step 9. Repeat Steps 4–8 B times.

Remark 1. Note that under the strict stationarity assumption, σ_t^2 can be uniquely represented in terms of past observations as

$$\sigma_t^2 = c_0 + \sum_{i=1}^{\infty} c_i Y_{t-i}^2, \tag{3.17}$$

where the c_i's depend on the GARCH parameters α_j and β_j, $j = 1, \ldots, m$ (Berkes et al., 2003). Thus, σ_t^2 is deterministic conditional on the past observations. As addressed in Pascaul et al. (2006), if the model parameters are known, the one-step ahead volatility is perfectly predictable given $\{Y_t\}_{t=1}^{n}$. The only uncertainty associated with the one-step ahead prediction comes from parameter estimation whose variability goes to 0 as $n \to \infty$.

Now, we can use the bootstrap distributions of Y_{n+h}^{2*} and σ_{n+h}^{2*} produced by Steps 4–9, i.e. $\hat{F}_{Y_{n+h}^{2*}}$ and $\hat{F}_{\sigma_{n+h}^{2*}}$ respectively, to approximate the unknown distributions of Y_{n+h}^2 and σ_{n+h}^2, for

$h = 1, \ldots, s$. Hence, the $100(1-\alpha)\%$ prediction interval of Y^2_{n+h} is given by

$$[0, H^*_{n+h}(1-\alpha)], \quad h = 1, \ldots, s, \tag{3.18}$$

where $H^*_{n+h}(1-\alpha)$ is the $(1-\alpha)$ quantile of $\hat{F}_{Y^{2*}_{n+h}}$. The corresponding $100(1-\alpha)\%$ prediction interval of Y_{n+h} (PI^*_y) is

$$[Q^*_{n+h}(\alpha/2), Q^*_{n+h}(1-\alpha/2)] \tag{3.19}$$

where $Q^*_{n+h}(\alpha/2) = -\sqrt{H^*_{n+h}(1-\alpha)}$ and $Q^*_{n+h}(1-\alpha/2) = \sqrt{H^*_{n+h}(1-\alpha)}$, assuming the underlying process is symmetrically distributed. Similarly, the $100(1-\alpha)\%$ bootstrap prediction interval of σ^2_{n+h} ($PI^*_{\sigma^2}$) is given by

$$[0, K^*_{n+h}(1-\alpha)], \tag{3.20}$$

where $K^*_{n+h}(1-\alpha)$ is the $(1-\alpha)$ quantile of $\hat{F}_{\sigma^{2*}_{n+h}}$.

Remark 2. Note that to save computing time, similar to the conditional bootstrap approaches of Cao et al. (1997) and Miguel and Olave (1999), we can omit the re-estimation Steps 4–6 in our algorithm. We call the simplified procedure conditional sieve bootstrap (CSB). However, CSB does not take into account the variance due to parameter estimation. Consequently, the prediction interval of the one-step ahead forecast of volatility is not computable.

3.2 Numerical Results

Here we investigate the performance of our method by simulations from various ARCH(p) and GARCH(p, q) models, with N$(0, 1)$ and t_5 error distributions. For every combination of model and error distribution, we compare our new unconditional sieve bootstrap (USB) and conditional sieve bootstrap (CSB) with the method proposed by Pascual et al. (2006) (PRR) based on the following algorithm:

- Step 1. Simulate the series and generate $R = 1000$ future values Y_{n+h} and σ^2_{n+h}, for $h = 1, \ldots, s$. Then, the empirical length of the prediction interval of Y_{n+h} is obtained by

$$L^e_{n+h,y} = Y^{R(1-\alpha/2)}_{n+h} - Y^{R(\alpha/2)}_{n+h}, \tag{3.21}$$

and that of σ^2_{n+h} is given by

$$L^e_{n+h,\sigma^2} = \sigma^{2,R(1-\alpha/2)}_{n+h} - \sigma^{2,R(\alpha/2)}_{n+h}. \tag{3.22}$$

- Step 2. Compute $B = 1000$ bootstrap forecasts $\{Y^{*,b}_{n+h}\}^B_{b=1}$ and $\{\sigma^{2*,b}_{n+h}\}^B_{b=1}$ and then construct the $100(1-\alpha)\%$ PI^*_y and $PI^*_{\sigma^2}$, $h = 1, \ldots, s$. The lengths of PI^*_y and $PI^*_{\sigma^2}$ are obtained respectively by

$$L^{*,M}_{n+h,y} = Q^*_{n+h}(1-\alpha/2) - Q^*_{n+h}(\alpha/2), \tag{3.23}$$

and

$$L^{*,M}_{n+h,\sigma^2} = K^*_{n+h}(1-\alpha). \tag{3.24}$$

- Step 3. Estimate the coverage of PI^*_y and $PI^*_{\sigma^2}$ using

$$C^{*,M}_{n+h,y} = \frac{1}{R}\sum_{r=1}^{R} \mathbf{1}_{\{Q^*_{n+h}(\alpha/2) \le Y^{*,r}_{n+h} \le Q^*_{n+h}(1-\alpha/2)\}}. \tag{3.25}$$

and

$$C^{*,M}_{n+h,\sigma^2} = \frac{1}{R}\sum_{r=1}^{R} \mathbf{1}_{\{0 \le \sigma^{2*,r}_{n+h} \le K^*_{n+h}(1-\alpha/2)\}}, \tag{3.26}$$

for $h = 1, \ldots, s$. Note that the coverage is defined as the proportion of future values lying within the prediction interval.

- Step 4. Repeat Steps 1–3 $MC = 1000$ times. Compute the average and the standard error of the coverage of PI_y^* by

$$\bar{C}^*_{n+h,y} = \sum_{M=1}^{MC} C^{*,M}_{n+h,y}/MC, \tag{3.27}$$

$$s.e(\bar{C}^*_{n+h,y}) = \left\{ \sum_{M=1}^{MC} (C^{*,M}_{n+h,y} - \bar{C}^*_{n+h,y})^2/MC \right\}^{\frac{1}{2}} \tag{3.28}$$

and those of the length of PI_y^* by

$$\bar{L}^*_{n+h,y} = \sum_{M=1}^{MC} L^{*,M}_{n+h,y}/MC \tag{3.29}$$

$$s.e(\bar{L}^*_{n+h,y}) = \left\{ \sum_{M=1}^{MC} (L^{*,M}_{n+h,y} - \bar{L}^*_{n+h,y})^2/MC \right\}^{\frac{1}{2}} \tag{3.30}$$

Similarly, the average and the standard error of the coverage of $PI^*_{\sigma^2}$ are given by

$$\bar{C}^*_{n+h,\sigma^2} = \sum_{M=1}^{MC} C^{*,M}_{n+h,\sigma^2}/MC, \tag{3.31}$$

$$s.e(\bar{C}^*_{n+h,\sigma^2}) = \left\{ \sum_{M=1}^{MC} (C^{*,M}_{n+h,\sigma^2} - \bar{C}^*_{n+h,\sigma^2})^2/MC \right\}^{\frac{1}{2}} \tag{3.32}$$

and those of the length of $PI^*_{\sigma^2}$ are given by

$$\bar{L}^*_{n+h,\sigma^2} = \sum_{M=1}^{MC} L^{*,M}_{n+h,\sigma^2}/MC \tag{3.33}$$

$$s.e(\bar{L}^*_{n+h,\sigma^2}) = \left\{ \sum_{M=1}^{MC} (L^{*,M}_{n+h,\sigma^2} - \bar{L}^*_{n+h,\sigma^2})^2/MC \right\}^{\frac{1}{2}} \tag{3.34}$$

We compare the performance of USB, CSB and PRR in terms of $\bar{C}^*_{n+h,y}$, $\bar{L}^*_{n+h,y}$, $\bar{C}^*_{n+h,\sigma^2}$, $\bar{L}^*_{n+h,\sigma^2}$ and CPU time, based on the following three models [3]:

[3] We also investigated the algorithm of Migual and Olave (1999). Their results are found to be equivalent to CSB in terms of coverage and sharpness. However, the required CPU time is substantially higher than that of CSB. Hence, we omit the method of Migual and Olave from further comparison.

- Model 1: ARCH(1)

$$Y_t = \sigma_t \epsilon_t, \tag{3.35}$$

$$\sigma_t^2 = 0.1 + 0.4 Y_{t-1}^2, \tag{3.36}$$

- Model 2: ARCH(2)

$$Y_t = \sigma_t \epsilon_t, \tag{3.37}$$

$$\sigma_t^2 = 0.1 + 0.2 Y_{t-1}^2 + 0.15 Y_{t-2}^2, \tag{3.38}$$

- Model 3: GARCH(1,1)

$$Y_t = \sigma_t \epsilon_t, \tag{3.39}$$

$$\sigma_t^2 = 0.05 + 0.1 Y_{t-1}^2 + 0.85 \sigma_{t-1}^2, \tag{3.40}$$

where ϵ_t follows either N$(0, 1)$ or t_5. In our study, we set the significance level α to 0.05, which corresponds to a 95% prediction interval (PI).

As shown in Tables 3.1–3.5, USB and CSB provide coverage close to the nominal significance level for PI_y^* , especially for small and moderate sample sizes, while for larger n all three methods perform similarly. In short term volatility forecasts, USB outperforms PRR for all ARCH models (see Tables 3.1–3.4) and both methods yield equivalent results for GARCH models (see Table 3.5). For longer term volatility forecasts, PRR has a slight edge over USB and CSB for small sample sizes. The performance of all three methods tends to be equivalent for larger samples. On comparing USB and CSB, typically USB is somewhat more precise than CSB across all samples, models and distributions. Note that the empirical lengths of prediction

intervals of returns and volatilities shown in Tables 3.1–3.5 are obtained using equations (3.21) and (3.22).

Lead time	Sample size	Method	Average coverage for return($s.e$)	Average length for return($s.e$)	Average coverage for volatility($s.e$)	Average length for volatility($s.e$)
1		Empirical	95%	1.54	95%	–
	300	PRR	94.53 (0.02)	1.53 (0.49)	92.50 (0.26)	0.20 (0.24)
		USB	94.75 (0.04)	1.56 (0.21)	92.00 (0.27)	0.33 (0.11)
		CSB	94.51 (0.05)	1.57 (0.22)	–	–
	1000	PRR	94.83 (0.01)	1.53 (0.43)	93.80 (0.24)	0.18(0.16)
		USB	94.97 (0.04)	1.56 (0.15)	94.20 (0.23)	0.35(0.07)
		CSB	94.89 (0.05)	1.57 (0.23)	–	–
	3000	PRR	94.92 (0.01)	1.54 (0.41)	92.70 (0.26)	0.18 (0.12)
		USB	95.14 (0.04)	1.55 (0.14)	95.00 (0.22)	0.35 (0.05)
		CSB	95.03 (0.04)	1.54 (0.12)	–	–
10		Empirical	95%	1.61	95%	0.34
	300	PRR	94.61 (0.02)	1.60 (0.14)	93.43 (0.05)	0.36 (0.11)
		USB	94.84 (0.02)	1.62 (0.16)	91.24 (0.06)	0.33 (0.11)
		CSB	94.78 (0.02)	1.61 (0.16)	91.33 (0.06)	0.33 (0.11)
	1000	PRR	94.75 (0.01)	1.60 (0.09)	94.43 (0.02)	0.36 (0.06)
		USB	95.09 (0.01)	1.63 (0.10)	93.55 (0.03)	0.35 (0.07)
		CSB	95.07 (0.01)	1.62 (0.11)	93.60 (0.03)	0.35 (0.08)
	3000	PRR	94.84 (0.01)	1.60 (0.07)	94.79 (0.02)	0.36 (0.04)
		USB	95.14 (0.01)	1.62 (0.08)	94.32 (0.02)	0.35 (0.05)
		CSB	95.19 (0.01)	1.62 (0.08)	94.30 (0.02)	0.35 (0.05)
20		Empirical	95%	1.62	95%	0.36
	300	PRR	94.54 (0.02)	1.59 (0.14)	93.36 (0.05)	0.36 (0.11)
		USB	94.81 (0.02)	1.62 (0.16)	91.27 (0.06)	0.33 (0.11)
		CSB	94.86 (0.02)	1.62 (0.17)	91.29 (0.06)	0.33 (0.11)
	1000	PRR	94.82 (0.01)	1.60 (0.09)	94.46 (0.02)	0.36 (0.06)
		USB	95.14 (0.01)	1.63 (0.10)	93.62 (0.03)	0.35 (0.07)
		CSB	95.12 (0.01)	1.63 (0.11)	93.56 (0.03)	0.35 (0.08)
	3000	PRR	94.92 (0.01)	1.60 (0.07)	94.79 (0.02)	0.36 (0.04)
		USB	95.12 (0.01)	1.62 (0.08)	94.30 (0.02)	0.35 (0.05)
		CSB	95.11 (0.01)	1.62 (0.07)	94.35 (0.02)	0.35 (0.05)

Table 3.1: Prediction intervals for returns and volatilities of an ARCH(1) process following Model 1 with N$(0, 1)$ innovations.

Lead time	Sample size	Method	Average coverage for return($s.e$)	Average length for return($s.e$)	Average coverage for volatility($s.e$)	Average length for volatility($s.e$)
1		Empirical	95%	1.55	95%	–
	300	PRR	94.56 (0.02)	1.53 (0.54)	85.80 (0.35)	0.22 (0.31)
		USB	94.75 (0.04)	1.55 (0.22)	90.60 (0.29)	0.33 (0.23)
		CSB	94.81 (0.04)	1.55 (0.28)	–	–
	1000	PRR	94.77 (0.01)	1.53 (0.65)	89.10 (0.31)	0.21 (0.65)
		USB	95.00 (0.04)	1.55 (0.19)	93.20 (0.25)	0.33 (0.20)
		CSB	94.90 (0.04)	1.54 (0.17)	–	–
	3000	PRR	94.92 (0.01)	1.54 (0.55)	91.70 (0.28)	0.19 (0.29)
		USB	95.08 (0.03)	1.54 (0.27)	95.00 (0.22)	0.33 (0.09)
		CSB	95.06 (0.03)	1.54 (0.15)	–	–
10		Empirical	95%	1.62	95%	0.35
	300	PRR	94.59 (0.02)	1.59 (0.21)	92.24 (0.07)	0.37 (0.20)
		USB	95.11 (0.02)	1.66 (0.30)	90.34 (0.08)	0.34 (0.28)
		CSB	95.03 (0.02)	1.65 (0.28)	89.20 (0.10)	0.32 (0.20)
	1000	PRR	94.83 (0.01)	1.60 (0.14)	94.16 (0.03)	0.36 (0.12)
		USB	95.35 (0.01)	1.65 (0.17)	92.58 (0.04)	0.33 (0.20)
		CSB	95.31 (0.01)	1.65 (0.17)	92.26 (0.04)	0.32 (0.12)
	3000	PRR	94.96 (0.01)	1.60 (0.10)	94.75 (0.02)	0.36 (0.07)
		USB	95.41 (0.01)	1.65 (0.12)	93.23 (0.03)	0.33 (0.09)
		CSB	95.44 (0.01)	1.65 (0.12)	93.34 (0.03)	0.33 (0.09)
20		Empirical	95%	1.64	95%	0.37
	300	PRR	94.56 (0.02)	1.59 (0.21)	92.16 (0.07)	0.37 (0.20)
		USB	95.10 (0.02)	1.66 (0.32)	90.23 (0.08)	0.34 (0.32)
		CSB	95.04 (0.02)	1.65 (0.29)	89.11 (0.10)	0.32 (0.22)
	1000	PRR	94.82 (0.01)	1.60 (0.14)	94.13 (0.03)	0.36 (0.12)
		USB	95.34 (0.01)	1.65 (0.18)	92.52 (0.04)	0.33 (0.20)
		CSB	95.26 (0.01)	1.65 (0.17)	92.16 (0.04)	0.32 (0.11)
	3000	PRR	94.93 (0.01)	1.60 (0.10)	94.65 (0.02)	0.36 (0.07)
		USB	95.42 (0.01)	1.65 (0.12)	93.27 (0.03)	0.33 (0.09)
		CSB	95.43 (0.01)	1.66 (0.12)	93.34 (0.03)	0.33 (0.09)

Table 3.2: Prediction intervals for returns and volatilities of an ARCH(1) process following Model 1 with t_5 innovations.

Lead time	Sample size	Method	Average coverage for return($s.e$)	Average length for return($s.e$)	Average coverage for volatility($s.e$)	Average length for volatility($s.e$)
1		Empirical	95%	1.52	95%	–
	300	PRR	94.39 (0.02)	1.50 (0.30)	90.30 (0.30)	0.19 (0.10)
		USB	94.59 (0.04)	1.52 (0.16)	91.80 (0.27)	0.28 (0.09)
		CSB	94.65 (0.04)	1.51 (0.15)	–	–
	1000	PRR	94.80 (0.01)	1.52 (0.31)	95.10 (0.22)	0.18 (0.10)
		USB	94.91 (0.03)	1.52 (0.12)	93.60 (0.24)	0.28 (0.05)
		CSB	94.94 (0.03)	1.52 (0.11)	–	–
	3000	PRR	94.85 (0.01)	1.51 (0.32)	93.20 (0.25)	0.17 (0.13)
		USB	94.88 (0.03)	1.52 (0.11)	94.10 (0.24)	0.28 (0.03)
		CSB	94.89 (0.03)	1.52 (0.10)	–	–
10		Empirical	95%	1.54	95%	0.25
	300	PRR	94.56 (0.02)	1.54 (0.13)	93.55 (0.06)	0.29 (0.08)
		USB	94.70 (0.02)	1.54 (0.14)	91.92 (0.07)	0.28 (0.09)
		CSB	94.71 (0.02)	1.54 (0.14)	91.15 (0.08)	0.27 (0.09)
	1000	PRR	94.77 (0.01)	1.54 (0.08)	94.30 (0.03)	0.28 (0.04)
		USB	94.86 (0.01)	1.55 (0.08)	93.97 (0.04)	0.28 (0.05)
		CSB	94.87 (0.01)	1.54 (0.08)	93.83 (0.04)	0.28 (0.05)
	3000	PRR	94.84 (0.01)	1.54 (0.07)	94.68 (0.02)	0.27 (0.03)
		USB	94.88 (0.01)	1.54 (0.06)	94.61 (0.02)	0.28 (0.03)
		CSB	94.85 (0.01)	1.54 (0.06)	94.51 (0.02)	0.28 (0.03)
20		Empirical	95%	1.55	95%	0.28
	300	PRR	94.59 (0.02)	1.54 (0.12)	93.59 (0.06)	0.29 (0.08)
		USB	94.65 (0.02)	1.54 (0.14)	91.88 (0.07)	0.28 (0.09)
		CSB	94.70 (0.02)	1.54 (0.14)	91.11 (0.08)	0.27 (0.09)
	1000	PRR	94.79 (0.01)	1.54 (0.08)	94.28 (0.03)	0.28 (0.04)
		USB	94.91 (0.01)	1.55 (0.08)	93.94 (0.04)	0.28 (0.05)
		CSB	94.85 (0.01)	1.54 (0.08)	93.88 (0.04)	0.28 (0.05)
	3000	PRR	94.89 (0.01)	1.55 (0.07)	94.66 (0.02)	0.27 (0.03)
		USB	94.90 (0.01)	1.55 (0.06)	94.66 (0.02)	0.28 (0.03)
		CSB	94.87 (0.01)	1.54 (0.06)	94.49 (0.02)	0.28 (0.03)

Table 3.3: Prediction intervals for returns and volatilities of an ARCH(2) process following Model 2 with N$(0, 1)$ innovations.

Lead time	Sample size	Method	Average coverage for return($s.e$)	Average length for return($s.e$)	Average coverage for volatility($s.e$)	Average length for volatility($s.e$)
1		Empirical	95%	1.52	95%	–
	300	PRR	94.56 (0.02)	1.53 (0.55)	87.20 (0.33)	0.23 (0.45)
		USB	94.86 (0.03)	1.55 (0.25)	90.90 (0.29)	0.30 (0.22)
		CSB	94.73 (0.03)	1.53 (0.26)	–	–
	1000	PRR	94.76 (0.01)	1.54 (0.39)	91.20 (0.28)	0.20 (0.25)
		USB	94.89 (0.03)	1.52 (0.15)	92.30 (0.27)	0.28 (0.10)
		CSB	94.77 (0.03)	1.52 (0.15)	–	–
	3000	PRR	94.90 (0.01)	1.52 (0.36)	93.20 (0.25)	0.18 (0.14)
		USB	95.03 (0.03)	1.54 (0.37)	94.50 (0.23)	0.29 (0.07)
		CSB	95.14 (0.03)	1.52 (0.12)	–	–
10		Empirical	95%	1.58	95%	0.28
	300	PRR	94.72 (0.02)	1.58 (0.25)	92.40 (0.08)	0.35 (0.52)
		USB	95.13 (0.02)	1.61 (0.25)	91.03 (0.08)	0.30 (0.19)
		CSB	94.91 (0.02)	1.59 (0.34)	88.91 (0.11)	0.30 (0.68)
	1000	PRR	94.84 (0.01)	1.56 (0.15)	94.22 (0.04)	0.32 (0.61)
		USB	95.18 (0.01)	1.59 (0.15)	92.54 (0.05)	0.28 (0.10)
		CSB	95.18 (0.01)	1.59 (0.15)	92.52 (0.05)	0.28 (0.10)
	3000	PRR	94.92 (0.01)	1.56 (0.09)	94.84 (0.02)	0.30 (0.05)
		USB	95.28 (0.01)	1.59 (0.12)	93.93 (0.03)	0.29 (0.10)
		CSB	95.32 (0.01)	1.59 (0.11)	93.78 (0.03)	0.29 (0.07)
20		Empirical	95%	1.56	95%	0.30
	300	PRR	94.70 (0.02)	1.57 (0.25)	92.33 (0.08)	0.35 (0.59)
		USB	95.06 (0.02)	1.61 (0.26)	90.96 (0.08)	0.30 (0.20)
		CSB	94.93 (0.02)	1.60 (0.52)	88.91 (0.11)	0.34 (2.05)
	1000	PRR	94.86 (0.01)	1.56 (0.16)	94.18 (0.04)	0.32 (0.50)
		USB	95.22 (0.01)	1.59 (0.15)	92.56 (0.05)	0.28 (0.10)
		CSB	95.22 (0.01)	1.60 (0.16)	92.48 (0.05)	0.28 (0.10)
	3000	PRR	94.97 (0.01)	1.56 (0.10)	94.84 (0.02)	0.30 (0.05)
		USB	95.33 (0.01)	1.60 (0.11)	93.95 (0.03)	0.29 (0.08)
		CSB	95.26 (0.01)	1.59 (0.11)	93.78 (0.03)	0.29 (0.08)

Table 3.4: Prediction intervals for returns and volatilities of an ARCH(2) process following Model 2 with t_5 innovations.

Lead time	Sample size	Method	Average coverage for return($s.e$)	Average length for return($s.e$)	Average coverage for volatility($s.e$)	Average length for volatility($s.e$)
1		Empirical	95%	3.81	95%	–
	500	PRR	94.61 (0.02)	3.79 (0.89)	91.50 (0.28)	1.21 (1.37)
		USB	94.76 (0.04)	3.88 (0.45)	91.00 (0.29)	1.38 (0.51)
		CSB	94.69 (0.04)	3.86 (0.46)	–	–
	1000	PRR	94.74 (0.01)	3.82 (0.90)	93.40 (0.25)	1.14 (0.66)
		USB	94.88 (0.03)	3.85 (0.38)	93.40 (0.25)	1.30 (0.44)
		CSB	94.84 (0.04)	3.85 (0.36)	–	–
	3000	PRR	94.87 (0.01)	3.81 (0.86)	94.70 (0.22)	1.07 (0.61)
		USB	94.99 (0.03)	3.86 (0.31)	94.60 (0.23)	1.30 (0.48)
		CSB	94.75 (0.04)	3.88 (0.33)	–	–
10		Empirical	95%	3.86	95%	1.66
	500	PRR	94.49 (0.02)	3.88 (0.88)	92.01 (0.08)	1.91 (7.31)
		USB	94.67 (0.03)	3.92 (0.43)	90.12 (0.11)	1.66 (0.64)
		CSB	94.52 (0.03)	3.89 (0.44)	88.29 (0.13)	1.59 (0.60)
	1000	PRR	94.74 (0.02)	3.90 (0.63)	93.36 (0.05)	1.70 (0.81)
		USB	94.78 (0.02)	3.89 (0.35)	92.09 (0.08)	1.65 (0.50)
		CSB	94.69 (0.03)	3.89 (0.35)	91.84 (0.09)	1.64 (0.55)
	3000	PRR	94.86 (0.01)	3.89 (0.61)	94.39 (0.03)	1.67 (0.73)
		USB	94.84 (0.02)	3.89 (0.26)	94.12 (0.05)	1.69 (0.44)
		CSB	94.68 (0.03)	3.92 (0.28)	93.99 (0.05)	1.71 (0.47)
20		Empirical	95%	3.92	95%	1.80
	500	PRR	94.32 (0.02)	3.92 (1.50)	91.14 (0.08)	2.44 (20.58)
		USB	94.55 (0.02)	3.92 (0.43)	89.29 (0.10)	1.72 (0.67)
		CSB	94.36 (0.02)	3.90 (0.45)	87.53 (0.12)	1.66 (0.64)
	1000	PRR	94.65 (0.02)	3.92 (0.50)	92.83 (0.05)	1.82 (0.77)
		USB	94.66 (0.02)	3.90 (0.35)	91.31 (0.07)	1.74 (0.55)
		CSB	94.61 (0.02)	3.90 (0.34)	91.26 (0.08)	1.74 (0.62)
	3000	PRR	94.81 (0.01)	3.93 (0.43)	94.24 (0.03)	1.81 (0.60)
		USB	94.73 (0.02)	3.91 (0.25)	93.70 (0.04)	1.80 (0.44)
		CSB	94.66 (0.02)	3.92 (0.24)	93.75 (0.04)	1.81 (0.40)

Table 3.5: Prediction intervals for returns and volatilities of a GARCH$(1, 1)$ process following Model 3 with N$(0, 1)$ innovations.

Now we compare the three methods in terms of CPU time[4]. Figure 3.1 presents the dynamic of the estimated CPU time for various sample sizes based on $B = 1000$ and $MC = 1000$. Note that our results on CPU time are the averages of 100 repetitions. As indicated by Figure 3.1, CSB and USB substantially outperform PRR. In particular, PRR requires 100 times as much CPU time as that of USB for small sample sizes. Remarkably, CSB provides the best performance with only relatively minor loss in terms of sharpness and coverage.

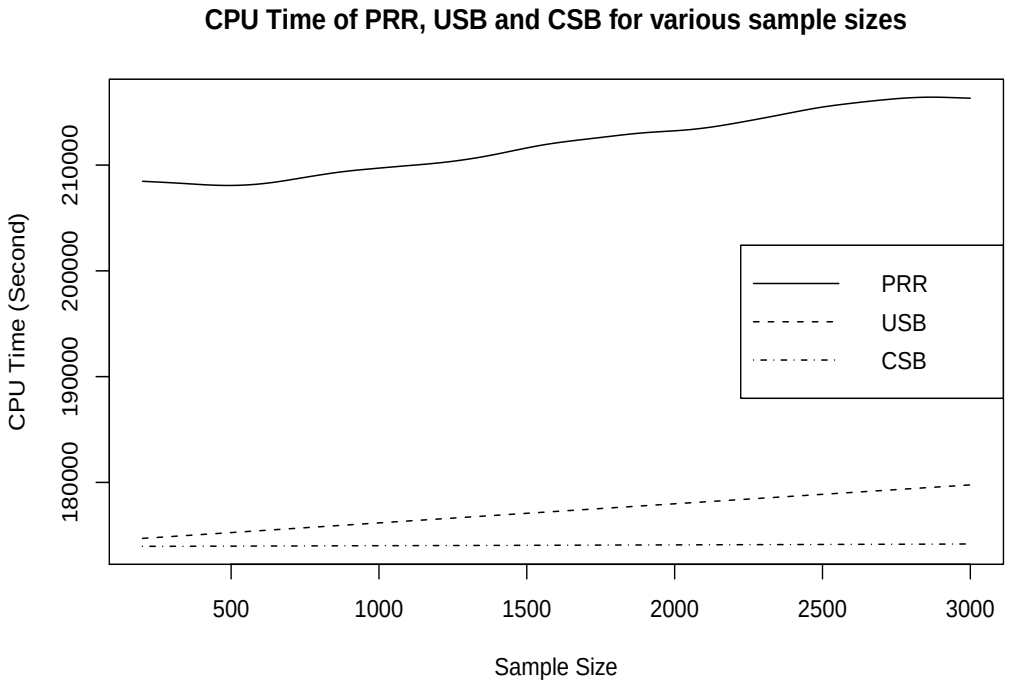

Figure 3.1: The estimated CPU time for PRR, USB and CSB applied to the GARCH(1,1) process of sample sizes from 200 to 3000.

[4]The computations were conducted on the Vidal cluster, which has 2 Operon processors with 4 GB RAM on each computing node.

Finally, USB and CSB typically yield some improvements in terms of returns while PRR generally provides slightly better results for volatilities. With increasing sample size, all three methods perform equivalently. However, USB and especially CSB are substantially less computationally demanding. Hence, USB and CSB may be selected as preferred procedures for constructing PIs for returns and volatilities. If only PIs for returns are of interest, then CSB is a better choice.

Remark 3. In practical applications, a fitted ARCH/GARCH model can exhibit a high degree of persistency. For example, for the case of an GARCH (1,1) model, high persistency means $\hat{\alpha}_1 + \hat{\beta}_1 \approx 1$, which may lead to instability of the LS estimation. Although in our studies we have not encountered any stability problems even for the cases of $\hat{\alpha}_1 + \hat{\beta}_1$ being 0.95 and 0.981, in practice it might lead to a failure of convergence and inflated standard errors, especially for small and moderate samples. Under these circumstances, we can follow the approach of Kristensen and Linton (2006) and censor the LS estimator at $1 - \epsilon$ for a small positive ϵ.

3.3 Case study

In this section, we apply the proposed USB and CSB algorithm to construct prediction intervals of returns and volatilities of the daily Yen/U.S Dollar exchange rate, i.e., the number of Yen per U.S Dollar. In order to avoid modeling particular weekend effects, we exclude all of the observations on Saturdays and Sundays (Anderson et al, 2003). Consequently, our full sample includes the daily average Yen/U.S exchange rate from March 28th, 1998 to July 28th, 2006, a total of 2175 observations (see Figure 3.2). Next, we transform the exchange rate into log returns using

$$Y_t = 100 * \log\left(\frac{\text{Yen/U.S exchange rate in day t}}{\text{Yen/U.S exchange rate in day t-1}}\right) \tag{3.41}$$

The new series of returns Y_t has a pattern shown in Figure 3.3, which is stationary and has mean close to zero.

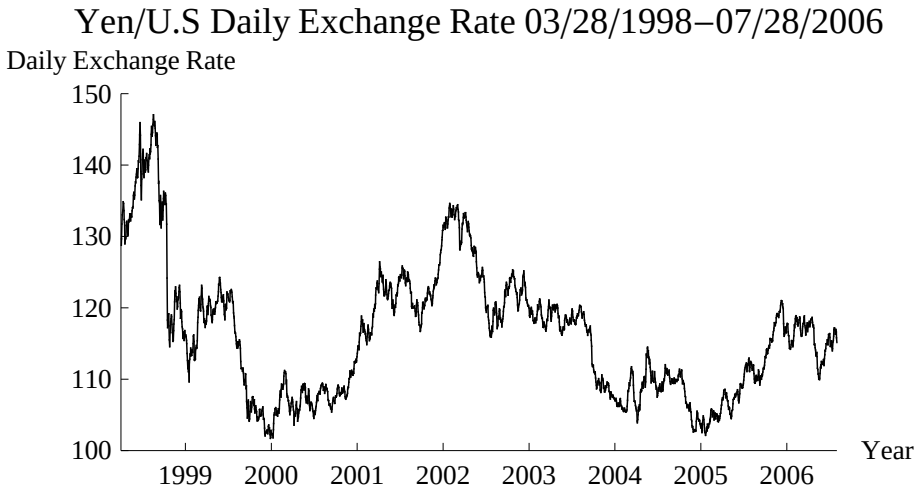

Figure 3.2: The Yen/U.S daily exchange rates from March 28th, 1998 to July 28th, 2006.

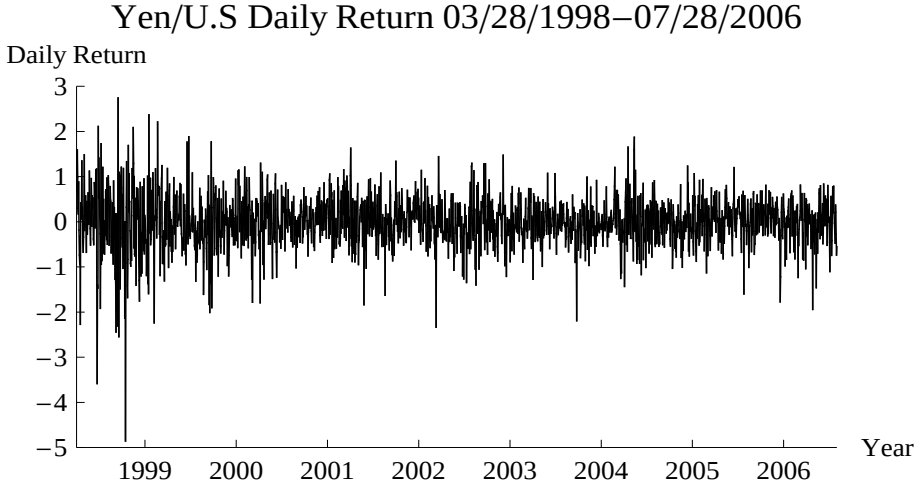

Figure 3.3: The Yen/U.S daily returns from March 28th, 1998 to July 28th, 2006.

Tables 3.6 and 3.7 present summary statistics of Y_t.

Mean	Median	S.D	Skewness	Kurtosis	Max.	Min.
-0.0024	0.0013	0.5611	-0.6037	8.0743	2.7413	-4.8567

Table 3.6: Summary statistics for log returns Y_t.

Autocorrelations	$\gamma(1)$	$\gamma(2)$	$\gamma(3)$	$\gamma(10)$	$\gamma(15)$	$\gamma(20)$
Y_t	0.2106	0.0086	-0.0573	0.0323	-0.0190	-0.0024
Y_t^2	0.0742	0.2204	0.1033	0.0729	0.0916	0.0492

Table 3.7: The autocorrelations of log returns Y_t at different lags.

As Table 3.6 shows, the estimated kurtosis is considerably higher than 3, indicating that the return distribution is leptokurtic. The p-values of the Jarque-Bera test (Jarque and Bera, 1980), the Robust Jarque-Bera test (Gel and Gastwirth, 2008) and the SJ test (Gel et al., 2007) are all less than 0.0001, so there is strong evidence to reject the hypothesis that Y_t is Gaussian. Also, autocorrelations of squared returns are large. As discussed by West and Cho (1995) as well as Anderson and Bollerslev (1998), a GARCH$(1,1)$ is a suitable model for Y_t.

Next, we partition the full sample into an in-sample estimation set from March 28th, 1998 to June 15th, 2006 and an out-of-sample verification set from June 16th, 2006 to July 28th, 2006. That is, based on a sample of 2143 observations, we make 31-step ahead predictions. By equation (5.22), we fit an ARMA$(1,1)$ model to Y_t^2 using LS. The resulting estimated model is given by

$$Y_t^2 = 0.006 + 0.9810 Y_{t-1}^2 + v_t - 0.9284 v_{t-1}, \tag{3.42}$$

i.e. $\hat{\alpha}_0 = 0.006$, $\hat{\alpha}_1 = 0.0525$ and $\hat{\beta}_1 = 0.9284$. Consistent with the previous literature, the estimate $\hat{\alpha}_1 + \hat{\beta}_1$ is close to unity.

Based on the fitted model (3.42), we first construct the 95% PIs of returns Y_{t+h} from June 16th to July 28th, 2006, using CSB and USB procedures respectively. Figure 3.4 presents the estimated distributions of 1-step-ahead and 10-step-ahead squared returns.

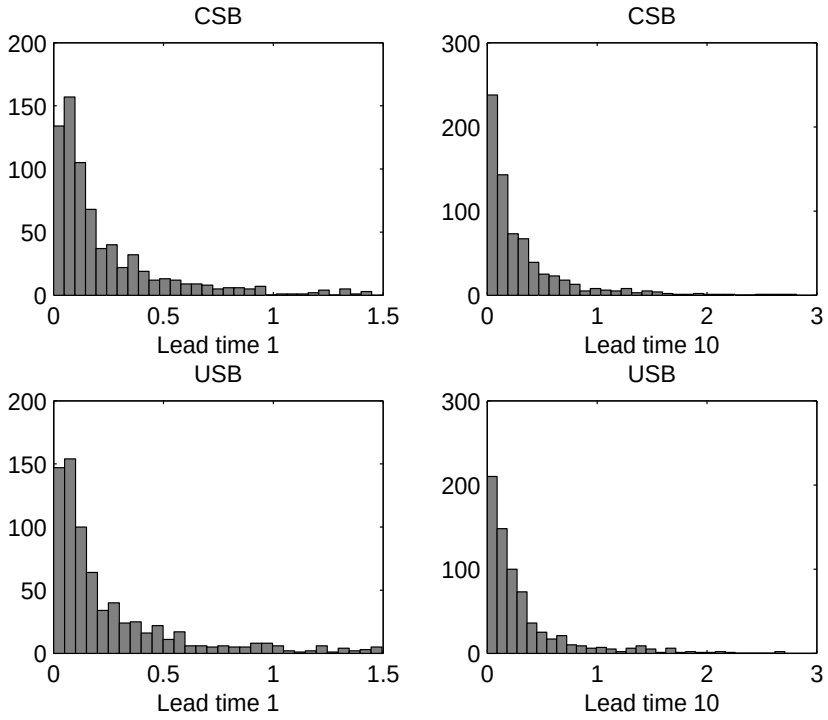

Figure 3.4: Histograms of bootstrap predictions of the future squared returns.

Note that since the squared returns are non-negative, each histogram shows a one-sided shape strictly greater than zero. From Figure 3.4, the estimated distributions obtained by CSB are very similar to those by USB. By taking the upper 95% quantiles of the estimated distributions, we

acquire the 95% PIs of returns Y_{t+h}, $h = 1, \ldots, 31$. Figures 4.4 and 3.6 show the 95% PIs of returns, provided by CSB and USB respectively, together with the true values of returns. Notice that the true observations are well covered by the PIs yielded by both CSB and USB.

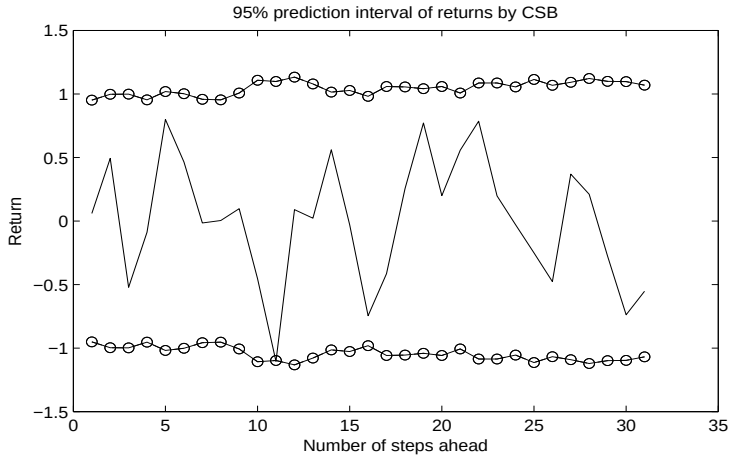

Figure 3.5: The 95% CSB prediction intervals of returns from June 16th to July 18th, 2006.

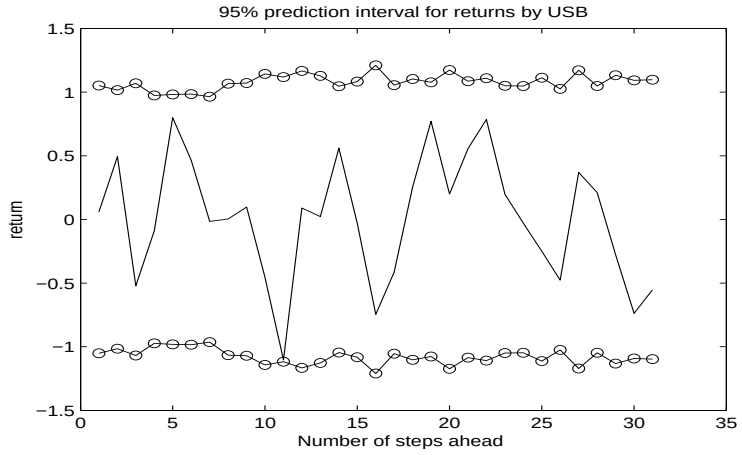

Figure 3.6: The 95% USB prediction intervals of returns from June 16th to July 18th, 2006.

Finally, we construct the 95% PIs for future volatilities of returns. Figure 3.7 shows the estimated distributions of future volatilities σ^2_{t+h}, $h = 1, \ldots, 31$.

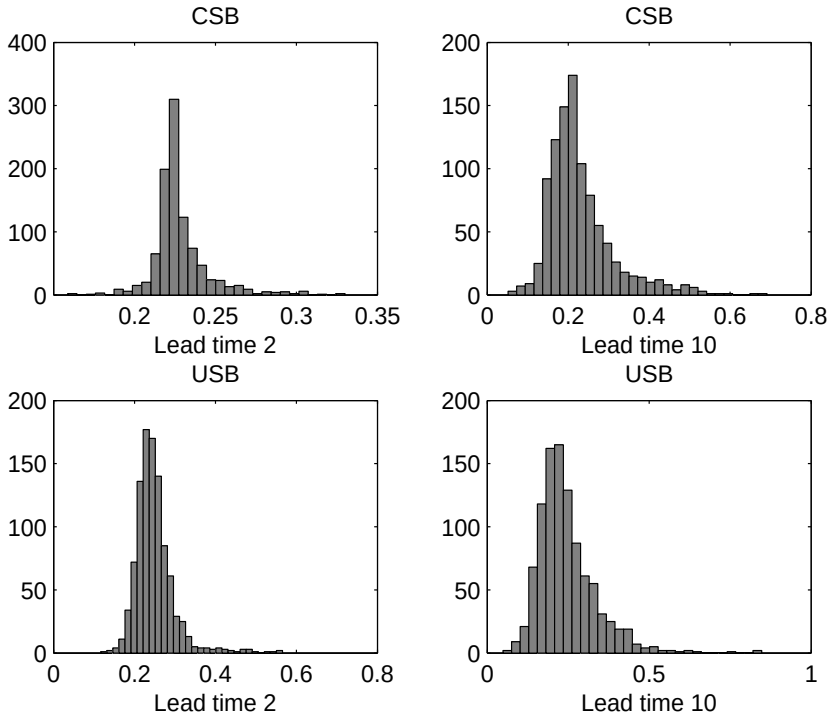

Figure 3.7: Histograms of bootstrap predictions of the future volatilities of returns.

Figure 3.9 indicates that the distributions of volatilities may be asymmetric. Similarly, we take the upper 95% quantiles of the estimated distributions to construct the PIs of volatilities using CSB and USB. In practice, we do not observe the volatility directly. For verification purposes, we calculate realized volatility from October 9th, 2005 to November 23rd, 2005 based on 5-minute returns using the following equation

$$\sigma_t^2 = Y_{t,1}^2 + \ldots + Y_{t,n}^2, \tag{3.43}$$

where n is the number of observations per day (Anderson and Bollerslev, 1998 and Taylor, 2005). Note that n is approximately 268 in our sample. Figures 3.8 and 3.9 present the 95% PIs for volatilities together with the realized volatilities.

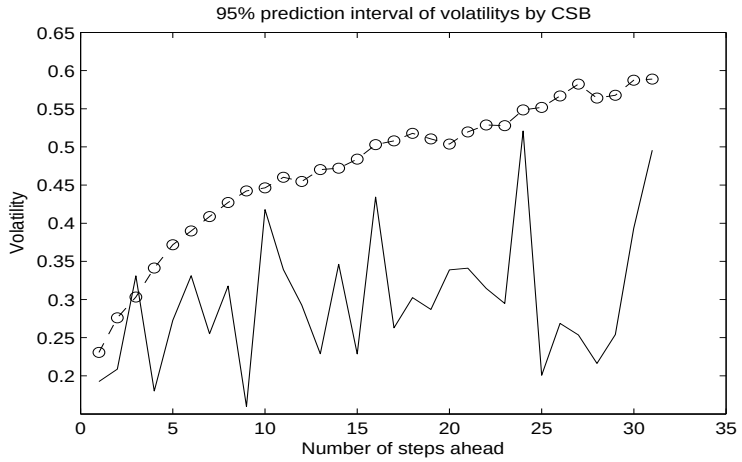

Figure 3.8: The 95% CSB prediction intervals of volatilities from June 16th to July 18th, 2006.

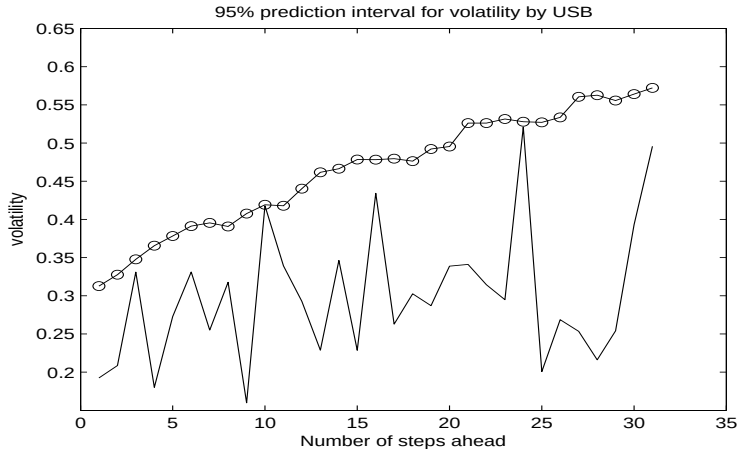

Figure 3.9: The 95% USB prediction intervals of volatilities from June 16th to July 18th, 2006.

In contrast to the PIs of returns, the 95% PIs of volatilities by USB outperform those of CSB. All of the realized volatilities lie within the 95% USB PIs, but one observation lies outside the boundary in the CSB case. Therefore, the consideration of variance due to parameter estimation is necessary if computational resources is plenty.

3.4 Discussion

In this chapter we propose a novel, fast and efficient method for constructing prediction intervals of returns and volatilities for ARCH/GARCH models. Our main idea is to transform the non-linear ARCH/GARCH re-sampling problem into a linear one by utilizing its AR/ARMA representation. Consequently, we can apply the sieve bootstrap procedure. The sieve bootstrap is an efficient residual-based re-sampling procedures for linear time series whose asymptotic and computational properties are well investigated. Adapting the sieve bootstrap in an ARCH/GARCH framework substantially decreases computational costs while providing competitively sharp and well calibrated prediction intervals for both returns and volatilities. The key reason for such improvement is that a linear AR/ARMA representation of ARCH/GARCH allows us to estimate all the model parameters using the recursive least squares (LS) or Yule-Walker (YW), which reduces the required computational time by up to 100 times, compared to other available re-sampling techniques for ARCH/GARCH models. As an alternative of the proposed sieve bootstrap procedure, it is interesting to apply blockwise bootstrap (Künsch, 1989; Liu and Singh, 1992) to construct PIs, which will be investigated in our future work.

The major disadvantage of the proposed method is that we lose the dependence structure of v_t on the corresponding Y_t's by bootstrapping. As a result, the bootstrap sample $\{Y_t^{2*}\}$ can be negative, which is explained by the following argument showing that v_t is unbounded.

Lemma 3.4.1

$$P(v_t < c) > 0, \ \textit{for any } c < 0. \tag{3.44}$$

Proof of Lemma 3.4.1 Let us start from a special case GARCH (1,1) model, where $\epsilon_t \sim N(0,1)$ and thus $\epsilon_t^2 \sim \chi_1^2$. We have

$$P(\epsilon_t^2 - 1 < -1/2) > 0. \tag{3.45}$$

Since

$$\sigma_t^2 = \alpha_0\bigg(1 + \sum_{l=1}^{\infty}\prod_{j=1}^{l}(\beta_1 + \alpha_1\epsilon_{t-j}^2)\bigg) \geq \alpha_0(1 + \beta_1 + \alpha_1\epsilon_{t-1}^2), \tag{3.46}$$

and

$$P\big(\alpha_0(1 + \beta_1 + \alpha_1\epsilon_{t-1}^2) > -2c\big) = P\bigg(\epsilon_t^2 > \frac{-2c - \alpha_0(1+\beta_1)}{\alpha_0\alpha_1}\bigg) > 0, \tag{3.47}$$

we obtain

$$\begin{aligned}
P(v_t < c) &= P\big((\epsilon_t^2 - 1)\sigma_t^2 < c\big) && (3.48)\\
&\geq P(\epsilon_t^2 - 1 < -1/2 \cap \sigma_t^2 > -2c)\\
&= P(\epsilon_t^2 - 1 < -1/2)P(\sigma_t^2 > -2c) && (3.49)\\
&\geq P(\epsilon_t^2 - 1 < -1/2)P\big(\alpha_0(1 + \beta_1 + \alpha_1\epsilon_{t-1}^2) > -2c\big) && (3.50)\\
&> 0.
\end{aligned}$$

Note that equation (3.49) is by independence and inequality (3.50) is by (3.46). Therefore, v_t is unbounded. The above arguments can be applied to any GARCH (p,q) model. In the GARCH (p,

q) case, equation (3.46) is generalized to

$$\begin{aligned}
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{\min(p,q)} (\beta_i + \alpha_i \epsilon_{t-i}^2)\sigma_{t-i}^2 + \delta_{p>q}(\alpha_{q+1}\sigma_{t-q-1}^2\epsilon_{t-q-1}^2 + \ldots + \alpha_p \sigma_{t-p}^2 \epsilon_{t-p}^2) \\
&+ \delta_{p<q}(\beta_{p+1}\sigma_{t-p-1}^2 + \ldots + \beta_q \sigma_{t-q}^2) \\
&\geq \alpha_0 + (\beta_1 + \alpha_1 \epsilon_{t-1}^2)\sigma_{t-1}^2 \\
&= \alpha_0 \left(1 + \sum_{l=1}^{\infty} \prod_{j=1}^{l} (\beta_1 + \alpha_1 \epsilon_{t-j}^2)\right) \\
&\geq \alpha_0 (1 + \beta_1 + \alpha_1 \epsilon_{t-1}^2).
\end{aligned}$$

Hence, the rest of the arguments still hold in the GARCH (p, q) model. □

Therefore, the bootstrap distributions of future returns $\{Y_{n+h}^{2*}\}$, hence the PIs, are not consistent for all quantiles in the same realization prediction, where h denotes the number of steps ahead. Let us first discuss the definitions of *same realization prediction* and *independent realization prediction* (Ing and Wei, 2005). Assume that $\{X_t\}_{t=1}^n$ is observed from an AR(p) process $\{X_t\}_{t\in\mathbb{Z}}$

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + e_t, \tag{3.51}$$

where $\{e_t\}$ is an i.i.d sequence with $Ee_t = 0$ and $Ee_t^2 < \infty$. Denote $\boldsymbol{a} = (a_1, \ldots, a_p)$ and $\hat{\boldsymbol{a}} = (\hat{a}_1, \ldots, \hat{a}_p)$, where $\hat{\boldsymbol{a}}$ is the LS estimates of $\boldsymbol{a}$ based on $\{X_t\}_{t=1}^n$. Now suppose $\{Z_t\}_{t=1}^n$ is a realization from an independent copy of $\{X_t\}_{t=1}^n$, i.e., $\{Z_t\}_{t=1}^n$ has exactly the same probabilistic structure as the $\{X_t\}_{t=1}^n$. The prediction of X_{n+h} is referred to as *same realization prediction* and that of Z_{n+h} is so-called *independent realization prediction*, in particular,

$$\hat{X}_{n+h} = \sum_{i=1}^{p} \hat{a}_i X_{n+h-i}, \tag{3.52}$$

and based on same $\hat{\boldsymbol{a}}$,

$$\hat{Z}_{n+h} = \sum_{i=1}^{p} \hat{a}_i Z_{n+h-i}. \tag{3.53}$$

Such concepts often appear in the context of forecasting. For practitioners, the emphasis is usually placed on same realization predictions. On the other hand, independent realization prediction is widely applied in analyzing structural properties due to its mathematical convenience. For example, AIC and its variants are proved to be asymptotically efficient for the purpose of predicting the future values of an independent realization. (Akaike, 1974; Shibata, 1980; Ing and Wei, 2005). For the proposed bootstrap procedure, the numerical results are promising, which potentially can be justified by showing

1. for *independent realization predictions*, the sieve bootstrap PIs are consistent.

2. for *same realization predictions*, the empirical tail distribution of Y_{n+h}^{2*} is a consistent estimate of $P\Big(Y_{n+h}^2 > F_{Y_{n+h}^2}^{-1}(\alpha)\Big)$, where $\alpha > C$ for some $C > 0$.

A possible remedy to our problem is: re-sampling from centered residuals $\{\hat{\epsilon}_{t,n}\}_{t=1}^n$ to obtain $\{\epsilon_t^*\}_{t=1}^n$ and then constructing $\{v_t^*\}$ by

$$v_t^* = (\epsilon_t^{*2} - 1)\hat{\sigma}_t^2, \;\; t = 1, \ldots, n. \tag{3.54}$$

As the dependence of v_t's originates from σ_t^2, this alternative bootstrap approach allows to keep such dependence and meanwhile to use the LS estimation. Next, let us investigate the distribution of $\{\epsilon_t^*\}$. Denote F as the cumulative distribution function of $\{\epsilon_t\}_{t\in\mathbb{Z}}$ and denote F_n as the empirical distribution of the unobserved residuals $\{\epsilon_t\}_{t=1}^n$. Also, define the empirical distribution function of the centered estimated residuals as

$$\hat{F}_n(x) = \frac{1}{n}\sum_{t=1}^{n} \mathbf{1}_{\{\hat{\epsilon}_{t,n} - \hat{\epsilon}_{\cdot,n} \leqslant x\}}, \tag{3.55}$$

where $\hat{\epsilon}_{\cdot,n} = n^{-1}\sum_{t=1}^n \hat{\epsilon}_{t,n}$. Note that $\{\epsilon_t^*\}_{t=1}^n$ i.i.d.$\sim \hat{F}_n$. In order to study $\hat{F}_n$, we introduce the definition of metric d_r, $r \geqslant 1$, for probability measure P and Q on $\mathbb{R}^k$ with $\int |x|^r dP < \infty$ and

$\int |x|^r dQ < \infty$:

$$d_r(P, Q) = \inf(E|X - Y|^r)^{1/r}, \tag{3.56}$$

where the infimum is taken over all pairs (X, Y) of random variables with X and Y distributed according to P and Q respectively. In particular, we consider the case when $r = 2$, so-called Mallows metric. Also, we impose the following additional assumptions:

1. $E\epsilon_t^4 < \infty$.

2. $\left(E\epsilon_t^4\right)^{1/2} \sum_{i=1}^p \alpha_i < 1$. Note that Giraitis et al.(2000) show that under this assumption, the ARCH equations have a unique strictly stationary solution such that $EY_t^4 < \infty$ and the squares Y_t^2 have a Volterra representation

$$Y_t^2 = \sum_{l=0}^{\infty} \sum_{i_1,\ldots,i_l}^{p} \alpha_{i_1} \ldots \alpha_{i_l} \epsilon_t^2 \epsilon_{t-i_1}^2 \ldots \epsilon_{t-i_1-\ldots-i_l}^2. \tag{3.57}$$

   That is, $Y_t^2 = f(\epsilon_t, \epsilon_{t-1}, \ldots)$ and thus the sequence $\{Y_t^2\}_{t\in\mathbb{Z}}$ is ergodic.

3. Let $\theta = (\alpha_0, \alpha_1, \ldots, \alpha_p)$ and $\hat{\theta}_n = (\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_p)$, which denote the unknown coefficients and their sample estimates respectively. We assume that the estimator $\hat{\theta}_n$ is asymptotically consistent with rate $\sqrt{n}$, i.e.,

$$\sqrt{n}\|\hat{\theta}_n - \theta\| = O_p(1). \tag{3.58}$$

   Note that both the maximum likelihood and the LS estimators satisfy assumption 3.

The next theorem states that the distribution of $\{\epsilon_t^*\}$ converges to that of $\{\epsilon_t\}$ in probability.

Theorem 3.4.2 *$d_2(F, \hat{F}_n) \to 0$ in probability as $n \to \infty$.*

Proof of Theorem 3.4.2 By Lemma 8.4 from Bickel and Freedman (1981), we have

$$d_2(F_n, F) \to 0, \quad \text{as } n \to \infty \text{ almost surely.} \tag{3.59}$$

Next, let J be uniformly distributed on $\{1, \ldots, n\}$, i.e., $J = j$ with probability $1/n$ for each of $j = 1, \ldots, n$. Define random variables X_1 and Y_1 with marginal distributions F_n and $\hat{F}_n$, respectively, according to

$$X_1 = \epsilon_J, \tag{3.60}$$

$$Y_1 = \hat{\epsilon}_{J,n} - \hat{\epsilon}_{\bullet,n}. \tag{3.61}$$

Then,

$$\begin{aligned} \{d_2(F_n, \hat{F}_n)\}^2 = \inf E(X - Y)^2 &\leqslant E_J(X_1 - Y_1)^2 && (3.62)\\ &= \frac{1}{n}\sum_{j=1}^{n}\left(\hat{\epsilon}_{j,n} - \epsilon_j - \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_{i,n}\right)^2 && (3.63)\\ &\leqslant \frac{6}{n}\sum_{j=1}^{n}(\hat{\epsilon}_{j,n} - \epsilon_j)^2 + \frac{3}{n^2}\left(\sum_{j=1}^{n}\epsilon_j\right)^2 \end{aligned}$$

Let

$$\zeta_{1,n} := \frac{1}{n}\sum_{j=1}^{n}(\hat{\epsilon}_{j,n} - \epsilon_j)^2, \tag{3.64}$$

$$\zeta_{2,n} := \frac{1}{n^2}\left(\sum_{j=1}^{n}\epsilon_j\right)^2. \tag{3.65}$$

By Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\epsilon_j = O_p(1), \tag{3.66}$$

and hence

$$\zeta_{2,n} = o_p(1). \tag{3.67}$$

Next, our goal is to show $\zeta_{1,n} = o_p(1)$. We adopt the notations in Horváth and Kokoszka (2001). Consider a $(p+1)$-dimensional vector $\mathbf{s} = (s_0, \ldots, s_p)$. Let

$$\gamma_{t,n}(\mathbf{s}) = n^{-1/2}\big(s_0 + s_1 Y_{t-1}^2 + \ldots + s_p Y_{t-p}^2\big), \tag{3.68}$$

and

$$\delta_{t,n}^2(\mathbf{s}) = \frac{\sigma_t^2}{\sigma_t^2 - \gamma_{t,n}(\mathbf{s})}. \tag{3.69}$$

Let $\mathbf{T} = [-T, T]^{p+1}$. If $n > (T/\min_{1\leqslant i\leqslant p}\alpha_i)^2$, then

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \ldots + \alpha_p Y_{t-p}^2 && (3.70)\\ &\geqslant \min_{1\leqslant i\leqslant p}\alpha_i\big(1 + Y_{t-1}^2 + \ldots + Y_{t-p}^2\big) && (3.71)\\ &> \gamma_{t,n}(\mathbf{s}) \end{aligned}$$

for all $\mathbf{s} \in \mathbf{T}$ and thus

$$\delta_{t,n}(\mathbf{s}) = \sqrt{\frac{\sigma_t^2}{\sigma_t^2 - \gamma_{t,n}(\mathbf{s})}}. \tag{3.72}$$

Also,

$$\hat{\epsilon}_{t,n} = \frac{Y_t}{\hat{\sigma}_{t,n}} = \epsilon_t \frac{\sigma_t}{\hat{\sigma}_{t,n}} = \epsilon_t \hat{\delta}_{t,n}, \tag{3.73}$$

where

$$\hat{\delta}_{t,n} = \frac{\sigma_t}{\hat{\sigma}_t} = \delta_{t,n}(n^{1/2}(\hat{\theta}_n - \theta)). \tag{3.74}$$

By (3.73),

$$\zeta_{1,n} = \frac{1}{n}\sum_{j=1}^{n} \epsilon_j^2(\hat{\delta}_{t,n} - 1)^2. \tag{3.75}$$

In order to show $\zeta_{1,n} = o_p(1)$, it is sufficient to prove that for any $T > 0$,

$$\frac{1}{n}\sup_{\mathbf{s}\in\mathbf{T}}\left|\sum_{j=1}^{n}\epsilon_j^2\left(\sqrt{\frac{\sigma_j^2}{\sigma_j^2 - \gamma_{j,n}(\mathbf{s})}} - 1\right)^2\right| = o_p(1). \tag{3.76}$$

Since

$$\begin{aligned}
\sup_{\mathbf{s}\in\mathbf{T}} \quad & \left|\frac{1}{n}\sum_{j=1}^{n}\epsilon_j^2\left(\sqrt{\frac{\sigma_j^2}{\sigma_j^2-\gamma_{j,n}(\mathbf{s})}}-1\right)^2\right| \qquad (3.77)\\
\leqslant \quad & \frac{1}{n}\sum_{j=1}^{n}\sup_{\mathbf{s}\in\mathbf{T}}\left|\epsilon_j^2\left(\sqrt{\frac{\sigma_j^2}{\sigma_j^2-\gamma_{j,n}(\mathbf{s})}}-1\right)^2\right|\\
\leqslant \quad & \frac{1}{n}\sum_{j=1}^{n}\left|\epsilon_j^2\left(\sqrt{\left\{1-\frac{T}{\sqrt{n}\sigma_j^2}\left(1+Y_{j-1}^2+\ldots+Y_{j-p}^2\right)\right\}^{-1}}-1\right)^2\right|\\
+ \quad & \frac{1}{n}\sum_{j=1}^{n}\left|\epsilon_j^2\left(\sqrt{\left\{1+\frac{T}{\sqrt{n}\sigma_j^2}\left(1+Y_{j-1}^2+\ldots+Y_{j-p}^2\right)\right\}^{-1}}-1\right)^2\right|
\end{aligned}$$

Let $C = T/\sigma_j^2\left(1+Y_{j-1}^2+\ldots+Y_{j-p}^2\right)$ and denote

$$\begin{aligned}
\zeta_{3,n} &:= \frac{1}{n}\sum_{j=1}^{n}\left|\epsilon_j^2\left(\sqrt{\left\{1-\frac{C}{\sqrt{n}}\right\}^{-1}}-1\right)^2\right|, \qquad (3.78)\\
\zeta_{4,n} &:= \frac{1}{n}\sum_{j=1}^{n}\left|\epsilon_j^2\left(\sqrt{\left\{1+\frac{C}{\sqrt{n}}\right\}^{-1}}-1\right)^2\right|.
\end{aligned}$$

Apply Taylor's expansion, we have

$$\begin{aligned}
\sqrt{\left\{1-\frac{C}{\sqrt{n}}\right\}^{-1}} &= 1+C/(2\sqrt{n})+O(1/n), \qquad (3.79)\\
\sqrt{\left\{1+\frac{C}{\sqrt{n}}\right\}^{-1}} &= 1-C/(2\sqrt{n})+O(1/n). \qquad (3.80)
\end{aligned}$$

Hence,

$$\zeta_{3,n} = o_p(1) \text{ and } \zeta_{4,n} = o_p(1), \qquad (3.81)$$

and therefore,

$$\frac{1}{n}\sup_{\mathbf{s}\in\mathbf{T}}\left|\sum_{j=1}^{n}\epsilon_j^2\left(\sqrt{\frac{\sigma_j^2}{\sigma_j^2-\gamma_{j,n}(\mathbf{s})}}-1\right)^2\right| = o_p(1), \qquad (3.82)$$

i.e., $\zeta_{1,n} = o_p(1)$, which completes the proof. $\square$

By Theorem 3.4.2, it follows immediately that 1-step-ahead forecast is consistent. We will further investigate and provide theoretical justification for multi-step-ahead forecasts in the context of same and independent realizations.

Chapter 4

Lagrange Multiplier Test for GARCH Effect Using Permutation/Bootstrap

One of the principal tools for modeling volatility dynamics in financial applications is the Autoregressive Conditional Heteroscedastic (ARCH) model (Engle, 1982), the Generalized ARCH (GARCH) model (Bollerslev, 1986) and their various extensions (e.g., Engle, 2002; Park, 2002; Conrad and Karanasos, 2006; Medeiros and Veiga, 2009). The first crucial step in the modeling procedure is to assess the presence of ARCH/GARCH effects, which is usually carried out by the Lagrange Multiplier (LM) test (Breusch and Pagan, 1979; Cook and Weisberg, 1983; Engle, 1982; Lee, 1991). The LM statistic asymptotically follows a χ^2-distribution and is simple and fast to calculate. Unsurprisingly, the LM test became quickly adopted in practice and nowadays is considered to be the major testing tool, being implemented in most software packages. However, numerous studies (Engle et al., 1985; Gregory, 1989; Peguin-Feissolle, 1999; Godfrey et al., 2004; Raunig, 2008) indicate that the LM test can be too conservative when data are non-Gaussian, especially for small and moderate samples. Hence, the LM test becomes unreliable in

many financial applications, where observations are usually skewed or heavy-tailed.

Several modifications are suggested to improve the robustness of the LM test under departures from normality. Koenker (1981) (see also Weiss, 1986) proposes a studentized version of the LM test statistic. However, such modification is not satisfactory for small sample sizes and the power of the test can be quite poor, except under Gaussian conditions. Dufour et al. (2004) suggests to employ the Monte Carlo (MC) technique to calculate the exact critical values of the LM test. The important precondition, however, is that the distribution of the data is known, which is usually not the case in practice. In addition, Godfrey et al. (2004) show that the MC test performs poorly when the assumed distribution is misspecified. Recently, Raunig (2008) proposes a two-stage rank-based test, i.e. first transforming the data to Gaussian by the Rosenblatt probability integral (Rosenblatt, 1952) and then applying a bootstrap procedure to obtain critical values; however, no asymptotic properties and theoretical justification are provided. In addition, the Raunig test is limited only to observations with absolutely continuous marginal distributions, which might be a quite restrictive assumption in financial applications (see Calvet and Fisher, 2008, and references therein).

In this chapter, we propose two novel, robust and data-driven procedures to improve size and power of the LM test, namely the permutation and bootstrap-based LM tests, which further extend the idea of the Koenker heteroscedasticity test for regression models (Koenker, 1981). Our simulation results indicate that both the permutation and bootstrap LM tests substantially outperform the parametric LM test, particularly in small and moderate samples under deviations from normality. In addition, we prove that the permutation LM (PLM) test is exact and the bootstrap LM (BLM) test is asymptotically correct. Hence, PLM may be viewed as the most preferable method for testing against conditional heteroscedasticity since the nominal level can be obtained in finite samples under minor assumptions on the observed data. On the other hand,

the BLM method that is only asymptotically correct and is generally less powerful than the PLM test, can be also extended beyond hypothesis testing, e.g. to sensitivity analysis, model selection procedures and assessment of prediction errors, under the ARCH/GARCH framework, i.e. to the cases when the alternative hypothesis of conditional heteroscedasticity is valid and, hence, the observed data are no longer exchangeable.

4.1 LM Test for ARCH/GARCH effects

Let us consider an autoregressive conditional heteroscedastic (ARCH) model:

$$\begin{aligned} Y_t &= \sigma_t \epsilon_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 Y_{t-1}^2 + \ldots + \alpha_p Y_{t-p}^2, \end{aligned} \tag{4.1}$$

where ϵ_t are independent and identically distributed (i.i.d.) random variables with $E\epsilon_t = 0$, $E\epsilon_t^2 = 1$ and $E\epsilon_t^4 < \infty$. Let $v_t = Y_t^2 - \sigma_t^2 = \sigma_t^2(\epsilon_t^2 - 1)$. Then the ARCH($p$) model (4.1) can be written in a linear autoregressive (AR) form

$$Y_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \ldots + \alpha_p Y_{t-p}^2 + v_t, \tag{4.2}$$

where v_t is white noise. Alternatively, denoting $Z_t = Y_t^2$, $Z_t > 0$ and w.l.g $Z_t = 0$, $t \leq 0$, we get

$$Z_t = \alpha_0 + \alpha_1 Z_{t-1} + \ldots + \alpha_p Z_{t-p} + v_t, \tag{4.3}$$

or in a matrix form

$$\boldsymbol{Z} = \boldsymbol{\alpha}_0 + \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{v}. \tag{4.4}$$

Here $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$, $\boldsymbol{\alpha}_0 = (\alpha_0, \ldots, \alpha_0)'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p)'$, $\boldsymbol{v} = (v_1, \ldots, v_n)'$ and $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n)'$, where $\boldsymbol{X}_t = (Z_{t-1}, \ldots, Z_{t-p})$, $1 \leq t \leq p$.

Our goal is to determine whether ARCH effects are present in the observed data $\{Y_t\}_{t=-\infty}^{\infty}$, i.e. to test the null hypothesis of no conditional heteroscedasticity

$$H_0 : \boldsymbol{\alpha} = \mathbf{0} \tag{4.5}$$

vs. the alternative hypothesis of ARCH presence

$$H_a : \boldsymbol{\alpha} \neq \mathbf{0}. \tag{4.6}$$

Remark 1. Note that one might be interested in assessing conditional heteroscedasticity in data that are potentially generated by a Generalized ARCH (GARCH) model

$$\begin{aligned} Y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2. \end{aligned} \tag{4.7}$$

However, in view of the discussion by Lee (1991), the test procedures for ARCH and GARCH effects are identical. Hence, our discussion focuses only on the ARCH case.

Given a sample of observations $Z_1, \ldots, Z_n$, let $\boldsymbol{U}_t = (Z_{t-1} - \bar{Z}_n, \ldots, Z_{t-p} - \bar{Z}_n)$, $\boldsymbol{U} = (\boldsymbol{U}_1, \ldots, \boldsymbol{U}_n)'$ and $\bar{\boldsymbol{Z}} = (\bar{Z}_n, \ldots, \bar{Z}_n)'$ be a $n \times 1$-vector, where $\bar{Z}_n = n^{-1} \sum_{t=1}^{n} Z_t$ is the sample mean of $\{Z_t\}$. Then the LM test statistic T_n (Koenker, 1981) is defined by

$$T_n = \frac{(\boldsymbol{Z} - \bar{\boldsymbol{Z}})' \boldsymbol{U} (\boldsymbol{U}'\boldsymbol{U})^{-1} \boldsymbol{U}' (\boldsymbol{Z} - \bar{\boldsymbol{Z}})}{n^{-1} (\boldsymbol{Z} - \bar{\boldsymbol{Z}})' (\boldsymbol{Z} - \bar{\boldsymbol{Z}})}. \tag{4.8}$$

Notice that T_n is equivalent to nR^2, where R^2 is the coefficient of determination of the AR equation (4.3). Next, we establish the asymptotic distribution of T_n under the null hypothesis of no conditional heteroscedasticity and the assumption that the generating noise ϵ_t is i.i.d. but is not necessarily Gaussian.

Remark 2. Although the result on weak convergence of T_n is widely applied (Breusch and Pagan, 1979; Koenker, 1981; Engle, 1982), to our knowledge, its systematic proof under the non-Gaussian distribution in a dependent setting is not in the literature.

Lemma 4.1.1 *Let T_n be defined in (4.8). Then under H_0, as $n \to \infty$, T_n converges in distribution, i.e.*

$$T_n \xrightarrow{d} \chi^2_p. \tag{4.9}$$

The proof of Lemma 4.1.1 is given in Appendix.

4.1.1 Permutation LM test (PLM)

In this Section we introduce the new permutation LM (PLM) test and derive its theoretical properties. In particular, we show that the PLM test can achieve the exact desired level of test in finite samples.

Let us start from defining the following notations due to Hoeffding (1952) and Romano (1989), i.e. let $\boldsymbol{G}$ be the group of permutations on $\mathbb{R}^n$, consisting of $n!$ elements, and $\boldsymbol{gZ}$ be one particular permutation of $\boldsymbol{Z}$, for every $\boldsymbol{g} \in \boldsymbol{G}$. Under H_0, the components of $\boldsymbol{Z}$ are i.i.d and, hence, are *exchangeable*, i.e., their joint distribution is invariant under permutations of components (Lehmann and Romano, 2005). Hence, $\boldsymbol{gZ}$ and $\boldsymbol{Z}$ have the same distribution for all $\boldsymbol{g} \in \boldsymbol{G}$.

The key idea of the PLM test is to estimate the unknown distribution of the LM test statistic T_n (see (4.8)) under H_0, by assigning an equal mass to the values $T_n(\boldsymbol{gZ})$, $\boldsymbol{g} \in \boldsymbol{G}$. In practice, computing this distribution is inefficient and often infeasible since the number of elements in $\boldsymbol{G}$ is generally very large. An alternative way is to compute an approximate estimate by sampling

without replacement $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_S$ from $\boldsymbol{G}$, where $S \leq n!$, and assigning an equal mass to $T_n(\boldsymbol{g}_i \boldsymbol{Z})$, $i = 1, \ldots, S$, which leads to the following procedure:

- **Step 1**. Choose $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_S$ without replacement from $\boldsymbol{G}$.
- **Step 2**. Given a sample $z = (Z_1, \ldots, Z_n)$, evaluate the permutation test statistic $T_n(\boldsymbol{g}_i z)$ for $i = 1, \ldots, S$.

The unknown distribution of T_n under H_0 is then approximated by

$$\hat{F}^p_{T_n}(x) = \sum_{i=1}^{S} \mathbf{1}_{\{T_n(\boldsymbol{g}_i z) < x\}} \tag{4.10}$$

and a permutation critical value at significance level α is given by the upper α-quantile of $\hat{F}^p_{T_n}(x)$, denoted by $Q^p_{T_n}(1-\alpha)$. The null hypothesis of no conditional heteroscedasticity is rejected if $T_n \geq Q^p_{T_n}(1-\alpha)$. If $S = n!$, then the resulting test is called a *permutation* test, otherwise it is called an *approximate permutation* test.

Next, we show that the approximate PLM test is *exact*. For any observation $z = (Z_1, \ldots, Z_n) \in \mathbb{R}^n$, let

$$T_n^{(1)}(z) \leq T_n^{(2)}(z) \leq \ldots \leq T_n^{(S)}(z) \tag{4.11}$$

be the ordered values of $T_n(\boldsymbol{g}_i z)$, $i = 1, \ldots, S$. For a given significance level α, let

$$k = S - \lfloor S\alpha \rfloor, \tag{4.12}$$

where $\lfloor S\alpha \rfloor$ denotes the largest integer less or equal to $S\alpha$. Let $S^+(z)$ and $S^0(z)$ be the numbers of values $T_n^{(i)}(z)$, $i = 1, \ldots, S$, which are greater than $T_n^{(k)}(z)$ and equal to $T_n^{(k)}(z)$, respectively, and let

$$a_S(z) = \frac{S\alpha - S^+(z)}{S^0(z)}. \tag{4.13}$$

Since $S^+(z) \leq S - k \leq S\alpha$ and $S^+(z) + S^0(z) \geq S - k + 1 > S\alpha$, we have $0 \leq a_S(z) < 1$. Let $\phi_S(z)$ be the test function defined by

$$\phi_S(z) = \begin{cases} 1 & \text{if } T_n(z) > T_n^{(k)}(z), \\ a_S(z) & \text{if } T_n(z) = T_n^{(k)}(z), \\ 0 & \text{if } T_n(z) < T_n^{(k)}(z). \end{cases}$$

Lemma 4.1.2 *Assume that $Ev_t^2 < \infty$. Under H_0, if $Z_1, \ldots, Z_n$ are exchangeable, then*

$$E\phi_S(\boldsymbol{Z}) = \alpha, \tag{4.14}$$

i.e, the permutation test is exact.

Proof of Lemma 4.1.2 First, consider $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_S$ to be fixed. Note that for every $z \in \mathbb{R}^n$,

$$\sum_{i=1}^{S} \phi_S(\boldsymbol{g}_i z) = S^+(z) + a_S(z)S^0(z) = S\alpha. \tag{4.15}$$

Since $\boldsymbol{g}_i\boldsymbol{Z}$ and $\boldsymbol{Z}$ are identically distributed for $i = 1, \ldots, S$, we obtain

$$S\alpha = E\sum_{i=1}^{S} \phi_S(\boldsymbol{g}_i\boldsymbol{Z}) = \sum_{i=1}^{S} E\phi_S(\boldsymbol{Z}) = SE\phi_S(\boldsymbol{Z}). \tag{4.16}$$

Hence, $E\phi_S(\boldsymbol{Z}) = \alpha$. Since this result holds for any choice of $\boldsymbol{g}_1, \ldots, \boldsymbol{g}_S$, we can conclude that the permutation test is *exact*. □

4.1.2 Bootstrap LM test (BLM)

As an alternative to permutation, we can consider another resampling technique, namely bootstrap, to obtain a distribution of the LM test statistic T_n in finite samples. We show that the bootstrap LM (BLM) test is consistent.

Notice that under H_0 of no conditional heteroscedasticity, the observed data $\{Y_t\}_{t=-\infty}^{\infty}$ follow the model:

$$Z_t = \alpha_0 + v_t, \quad t = 1, \ldots, n, \tag{4.17}$$

where $v_t = \alpha_0(\epsilon_t^2 - 1)$ are i.i.d. such that $Ev_t = 0$ and $Ev_t^2 = \alpha_0^2 E(\epsilon_t^2 - 1)^2 < \infty$. Denote $\tau^2 = Ev_t^2$. Now we proceed with the following algorithm to obtain the bootstrap critical value of T_n:

- **Step 1**. Estimate the coefficient α_0 in equation (4.17) by the least squares (LS),

$$\hat{\alpha}_0 = \bar{Z}_n. \tag{4.18}$$

- **Step 2**. Calculate the estimated residuals $\{\hat{v}_t\}_{t=1}^n$ by

$$\hat{v}_t = Z_t - \hat{\alpha}_0. \tag{4.19}$$

  Note that since $n^{-1}\sum_{t=1}^n \hat{v}_t = 0$, we do not need to center the residuals. The empirical distribution of $\{\hat{v}_t\}_{t=1}^n$ is

$$\hat{F}_{v,n}(s) = \sum_{t=1}^{n} \mathbf{1}_{\{\hat{v}_t < s\}}. \tag{4.20}$$

- **Step 3**. Obtain the bootstrap residuals $\{v_t^*\}_{t=1}^n$ by sampling with replacement from $\hat{F}_{v,n}(s)$.

- **Step 4**. Construct the bootstrap sample $\{Z_t^*\}_{t=1}^n$ by

$$Z_t^* = \hat{\alpha}_0 + v_t^*. \tag{4.21}$$

- **Step 5**. Denote the bootstrap quantities by $\boldsymbol{Z}^* = (Z_1^*, \ldots, Z_n^*)'$, $\boldsymbol{U}_t^* = (Z_{t-1}^* - \bar{Z}_n^*, \ldots, Z_{t-p}^* - \bar{Z}_n^*)$, $\boldsymbol{U}^* = (\boldsymbol{U}_1^*, \ldots, \boldsymbol{U}_n^*)'$ and $\bar{\boldsymbol{Z}}^* = (\bar{Z}_n^*, \ldots, \bar{Z}_n^*)'$, where $\bar{Z}_n^* = n^{-1}\sum_{t=1}^n Z_t^*$. Calculate the bootstrap LM test statistic by

$$T_n^* = \frac{(\boldsymbol{Z}^* - \bar{\boldsymbol{Z}}^*)'\boldsymbol{U}^*(\boldsymbol{U}^{*'}\boldsymbol{U}^*)^{-1}\boldsymbol{U}^{*'}(\boldsymbol{Z}^* - \bar{\boldsymbol{Z}}^*)}{n^{-1}(\boldsymbol{Z}^* - \bar{\boldsymbol{Z}}^*)'(\boldsymbol{Z}^* - \bar{\boldsymbol{Z}}^*)} \tag{4.22}$$

- **Step 6**. Repeat **Steps 3–6** B times to obtain $\{T^*_{n,1}, \ldots, T^*_{n,b}, \ldots, T^*_{n,B}\}$, where B is the number of bootstrap replications and b denotes the b^{th} iteration.

The unknown distribution of T_n under H_0 is approximated by the bootstrap distribution of T^*_n:

$$\hat{F}^*_{T_n}(x) = \sum_{b=1}^{B} \mathbf{1}_{\{T^*_{n,b} < x\}}. \tag{4.23}$$

The bootstrap critical value at significance level α is obtained as the upper α-quantile of $\hat{F}^*_{T_n}(x)$, denoted by $Q^*_{T_n}(1-\alpha)$. The null hypothesis of no ARCH effects is rejected if $T_n \geq Q^*_{T_n}(1-\alpha)$.

Next, we show that the BLM test is *asymptotically correct*, i.e., it gives correct α-level under H_0, as $n \to \infty$.

Theorem 4.1.3 *If $Ev_t^4 < \infty$, then under H_0*

$$\sup_x |P^*(T^*_n \leq x) - P(T_n \leq x)| = o_p(1). \tag{4.24}$$

The proof of Theorem 4.1.3 is given in Appendix.

Remark 3. Note that the moment assumption required for the consistency of the BLM test is $Ev_t^4 < \infty$, i.e., $E\epsilon_t^8 < \infty$, which seems to be restrictive. However, our numerical studies show strong evidence that the BLM test can preserve the correct size of the test even when the moment assumptions are not satisfied, e.g., when ϵ_t follow the Student t-distribution with ν degrees of freedom and $0 < \nu \leq 8$. Figure 4.1 demonstrates size and power of the parametric, bootstrap and permutation LM tests, under Student t-distributions with 2, 3 and 4 degrees of freedom respectively and a significance level α of 0.05. Notice that although BLM does not perform as well as PLM, it significantly outperforms the parametric LM test in all cases. Remarkably, even under t_2-distribution where $E\epsilon_t^2 = \infty$, the size of the BLM test still shows a pattern of

convergence to the nominal value $\alpha = 0.05$ while the size of the parametric LM test fluctuates around 0.02 as sample size increases.

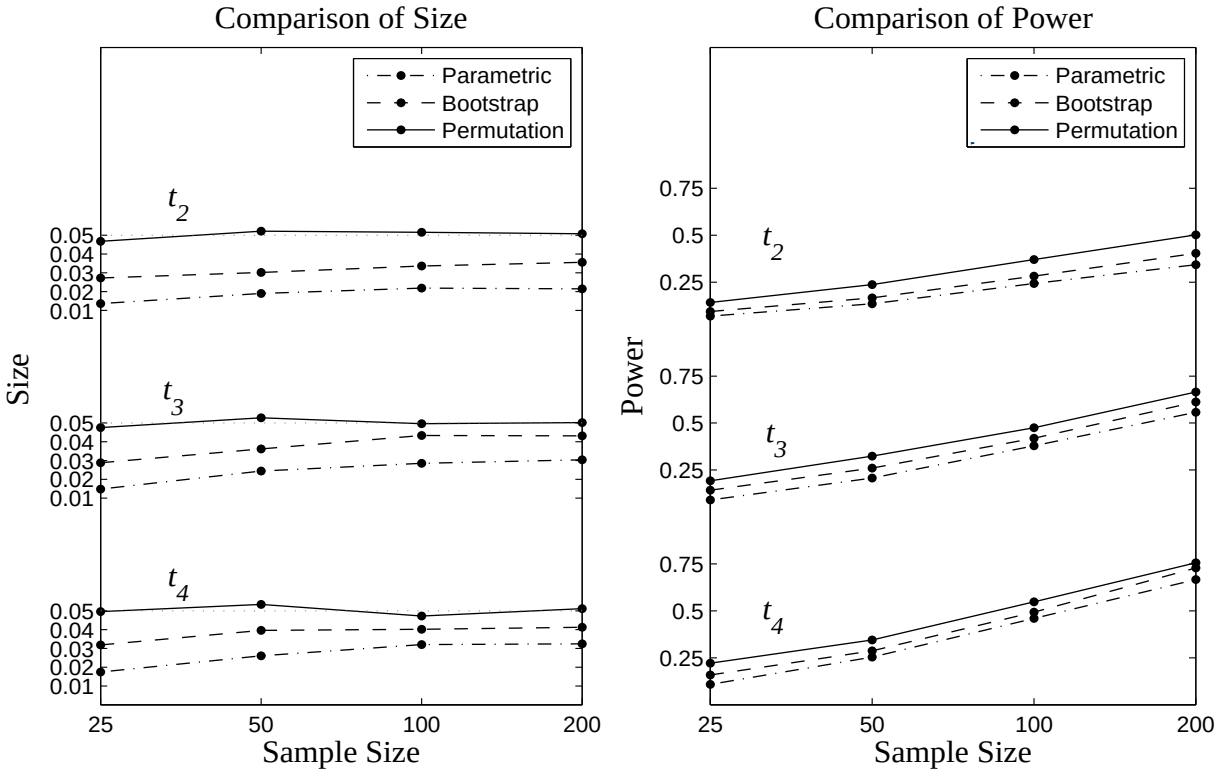

Figure 4.1: Size (left panel) and power (right panel) of the parametric, bootstrap and permutation LM tests under t_2, t_3 and t_4 distributions and a significance level $\alpha = 0.05$. For a power study the data are generated from ARCH(1). Number of Monte Carlo simulations is 5000 and number of bootstrap replications is 1000. The grey dashed line on the left panel represents a declared significance level α of 0.05.

4.2 Finite sample properties

In this Section we present a simulation study on size and power of the proposed permutation and bootstrap LM tests. In particular, we consider ARCH/GARCH models[1] under various skewed and heavy-tailed innovations, including a Student t-distribution with 5 degrees of freedom, χ^2-distribution with 2 degrees of freedom, Laplace distribution, Normal Inverse Gaussian (NIG) distributions[2] with different combinations of parameters and Tukey Contaminated Normal (CN) distributions [3]. For a study on size of the test, we compare the PLM and BLM tests to the parametric LM test with critical values approximated by a χ^2_p-distribution. For a power study, we evaluate performance of PLM and BLM with respect to the parametric LM test based on each of the χ^2_p-approximated critical values and the Monte Carlo simulated critical values, obtained under assumption that the underlying distribution of observations is known.

Table 4.1 shows sizes of the parametric, permutation and bootstrap LM tests respectively, at the significance level $\alpha = 0.05$. In general, estimated Type I errors of the PLM and BLM tests

[1]Due to limitation of space, here we present only numerical studies for detecting conditional heteroscedasticity effects for ARCH(1) and GARCH(1,1) models. However, the results are similar for higher order ARCH/GARCH models and are available upon request.

[2]The *NIG* distribution is solely determined by its four parameters α, β, μ and δ, corresponding to shape, skewness, location and scale. The mean, variance, skewness and kurtosis of *NIG* are respectively defined by $\mu + \beta\delta/\gamma$, $\delta\alpha^2/\gamma^3$, $3\beta/\alpha\sqrt{\delta\gamma}$ and $3(1 + 4\beta^2/\alpha^2)/\delta\gamma$, where $\gamma = \sqrt{\alpha^2 - \beta^2}$. Various combinations of these four parameters covers a broad range of continuous distribution of different shapes. Due to its flexibility, *NIG* is often used in modeling heavy-tailed financial processes (for more detailed discussion see, for example, Barndorff-Nielsen, 1997 and references therein).

[3]The Tukey family of *CN* distribution has the cumulative distribution function (cdf) $F_{CN_{\lambda,\alpha}}(x) = (1-\alpha)\Phi(x) + \alpha\Phi(x/\lambda)$, where $\Phi(x)$ is the cdf of standard normal distribution, $0 \leq \alpha \leq 1$ and $\lambda > 0$ are constants (Tukey, 1960; Gleason, 1993)

are consistently closer to the nominal level of 0.05 compared to the result yielded by the parametric LM test, for all distributions and sample sizes. The improvement in estimated Type I error is particularly significant in small and moderate samples from heavy-tailed and skewed distributions. For example, for 25 observations from a t_5-distribution (see Table 4.1), sizes yielded by the new PLM and BLM tests are 0.0488 and 0.0346 respectively, compared to 0.0272 provided by the parametric LM test; hence, the difference between the size of the test provided by PLM and a nominal α of 0.05 is 0.0012, which is almost 20 times smaller than the respective difference yielded by the parametric test. As expected, in general performance of all tests improves with an increase of a sample size. However, in contrast to PLM and BLM, the parametric LM test still noticeably deviates from the nominal level $\alpha = 0.05$ even for samples of 100 observations, especially for heavier tailed and skewed distributions. Notice that even for all samples from a normal distribution, the new PLM and BLM tests perform better than the parametric LM test, particularly for small sample sizes.

Distribution	Skewness	Kurtosis	Test	Sample Size			
				25	50	100	200
$N(0,1)$	0	3	Parametric	0.0268	0.0308	0.0426	0.0454
			Permutation	0.0510	0.0522	0.0488	0.0460
			Bootstrap	0.0416	0.0430	0.0480	0.0520
t_5	0	6	Parametric	0.0272	0.0300	0.0306	0.0346
			Permutation	0.0488	0.0490	0.0490	0.0492
			Bootstrap	0.0346	0.0384	0.0492	0.0490
χ^2_2	2	6	Parametric	0.0296	0.0314	0.0340	0.0416
			Permutation	0.0514	0.0506	0.0530	0.0526
			Bootstrap	0.0302	0.0304	0.0428	0.0446
Laplace	0	6	Parametric	0.0292	0.0236	0.0308	0.0362
			Permutation	0.0502	0.0460	0.0514	0.0538
			Bootstrap	0.0360	0.0372	0.0452	0.0500
NIG_1	0	12.5	Parametric	0.0293	0.0300	0.0322	0.0334
$\alpha = 0.4, \beta = 0$			Bootstrap	0.0346	0.0382	0.0444	0.0446
$\mu = 1$, $\delta = 0.6$			Permutation	0.0528	0.0462	0.0530	0.0514
NIG_2	2.5	21.5	Parametric	0.0298	0.0302	0.0298	0.0382
$\alpha = 0.5, \beta = 0.2$			Bootstrap	0.0306	0.0360	0.0410	0.0420
$\alpha = 0.5, \beta = 0.2$			Permutation	0.0498	0.0502	0.0534	0.0526
$CN_{0.25,0.4}$	0	6.4	Parametric	0.0276	0.0328	0.0398	0.0434
			Permutation	0.0498	0.0524	0.0502	0.0522
			Bootstrap	0.0388	0.0472	0.0498	0.0532

Table 4.1: Size of the bootstrap and permutation LM tests and the parametric LM test with χ^2_p-approximated critical values, under a nominal level $\alpha = 0.05$. Number of Monte Carlo simulations is 5000 and number of bootstrap/permutation replications is 999.

Tables 4.2 and 4.3 present analysis of power, yielded by the new PLM and BLM tests as well as the parametric LM test with χ^2_p-approximated critical values and the parametric LM test with Monte Carlo simulated critical values which are obtained under the assumption that the underlying distribution of data is known. Among all the considered tests, PLM shows the most superior performance, across all samples and distributions, especially for heavy tailed and skewed data of small and moderate sample sizes. BLM is generally less efficient in detecting conditional heteroscedasticity than the parametric LM test with Monte Carlo simulated critical values. However, one can employ the Monte Carlo simulated critical values only if the underlying distribution for observations is known (Dufour et al., 2004), which typically is not the case in practice. Moreover, as Godfrey et al. (2004) indicates, the parametric LM test with Monte Carlo simulated critical values performs poorly if the underlying distribution is misspecified. Hence, the new PLM and BLM offer powerful and robust approaches for detecting conditional heteroscedasticity, under no or limited knowledge on the observed data.

In short, due to its exactness, the PLM test provides the best performance in terms of size and power and may be viewed as the most preferable test for ARCH/GARCH effects; while BLM and its various potential modifications offer a data-driven methodology beyond hypothesis testing and may be employed, for example, for sensitivity analysis, model selection and assessment of prediction errors under an ARCH/GARCH framework.

Distribution	Skewness	Kurtosis	Test	Sample Size			
				25	50	100	200
$N(0,1)$	0	3	Parametric	0.1398	0.3458	0.6658	0.9152
			Permutation	0.2500	0.4446	0.8692	0.9298
			Bootstrap	0.1978	0.4092	0.7042	0.9290
t_5	0	6	Parametric	0.1128	0.2846	0.5008	0.7362
			Permutation	0.3238	0.5350	0.7726	0.9406
			Bootstrap	0.2648	0.4872	0.7404	0.9258
χ^2_2	2	6	Parametric	0.0916	0.2262	0.4342	0.6712
			Permutation	0.2170	0.3366	0.5472	0.7564
			Bootstrap	0.1342	0.2808	0.4952	0.7308
Laplace	0	6	Parametric	0.1016	0.2486	0.4642	0.7184
			Permutation	0.2112	0.3470	0.5666	0.7662
			Bootstrap	0.1518	0.2966	0.5152	0.7628
NIG_1	0	12.5	Parametric	0.1018	0.2108	0.3776	0.5924
$\alpha = 0.4, \beta = 0$			Permutation	0.1922	0.3026	0.4832	0.6916
$\mu = 1, \delta = 0.6$			Bootstrap	0.1420	0.2606	0.4364	0.6524
NIG_2	2.5	21.5	Parametric	0.1054	0.2000	0.3646	0.5642
$\alpha = 0.5, \beta = 0.2$			Permutation	0.1934	0.2996	0.4818	0.6698
$\mu = 1, \delta = 0.5$			Bootstrap	0.1412	0.2520	0.4270	0.6260
$CN_{0.25,0.4}$	0	6.4	Parametric	0.1794	0.3766	0.6716	0.9162
			Permutation	0.2502	0.4492	0.7136	0.9284
			Bootstrap	0.2184	0.4234	0.7036	0.9262

Table 4.2: Power of the bootstrap and permutation LM tests and the parametric LM test with the χ^2_p-approximated critical value for ARCH(1) processes with α_0 of 0.6 and α_1 of 0.4. Nominal level is $\alpha = 0.05$. Number of Monte Carlo simulations is 5000 and number of bootstrap/permutation replications is 999.

Distribution	Skewness	Kurtosis	Test	Sample Size			
				25	50	100	200
$N(0,1)$	0	3	Parametric	0.1766	0.3736	0.6966	0.9226
			Permutation	0.2472	0.4480	0.7288	0.9346
			Bootstrap	0.2112	0.4166	0.7198	0.9448
t_5	0	6	Parametric	0.1318	0.2828	0.5128	0.7600
			Permutation	0.2070	0.3644	0.6036	0.8162
			Bootstrap	0.1722	0.3230	0.5680	0.7998
χ^2_2	2	6	Parametric	0.1238	0.2420	0.4480	0.6998
			Permutation	0.2184	0.3452	0.5768	0.7784
			Bootstrap	0.1404	0.2930	0.5176	0.7606
Laplace	0	6	Parametric	0.1120	0.2516	0.4698	0.7324
			Permutation	0.2090	0.3362	0.5700	0.7872
			Bootstrap	0.1542	0.2952	0.5214	0.7760
NIG_1	0	12.5	Parametric	0.1070	0.2110	0.3868	0.6076
$\alpha = 0.4, \beta = 0$			Permutation	0.1896	0.3036	0.4884	0.7050
$\mu = 1, \delta = 0.6$			Bootstrap	0.1444	0.2620	0.4498	0.6660
NIG_2	2.5	21.5	Parametric	0.1054	0.2016	0.3736	0.5784
$\alpha = 0.5, \beta = 0.2$			Permutation	0.1884	0.2992	0.4854	0.6882
$\mu = 1, \delta = 0.5$			Bootstrap	0.1416	0.2530	0.4316	0.6416
$CN_{0.25,0.4}$	0	6.4	Parametric	0.1800	0.3796	0.6928	0.9188
			Permutation	0.2464	0.4464	0.7336	0.9468
			Bootstrap	0.2154	0.4332	0.7206	0.9310

Table 4.3: Power of the bootstrap and permutation LM tests and the parametric LM test with the χ^2_p-approximated critical value for GARCH(1,1) processes with α_0 of 0.6, α_1 of 0.4 and β_1 of 0.3. Nominal level is $\alpha = 0.05$. Number of Monte Carlo simulations is 5000 and number of bootstrap/permutation replications is 999.

4.3 Case studies

Let us consider two daily average exchange rate processes: the Yen/U.S. Dollar and the Euro/U.S. Dollar[4]. Similar to the case study of the previous chapter, we exclude all of the observations on Saturdays and Sundays in order to avoid modeling particular weekend effects. Consequently, each sample includes from April 20th, 1998 to July 28th, 2006, a total of 2159 observations (see Figure 4.2).

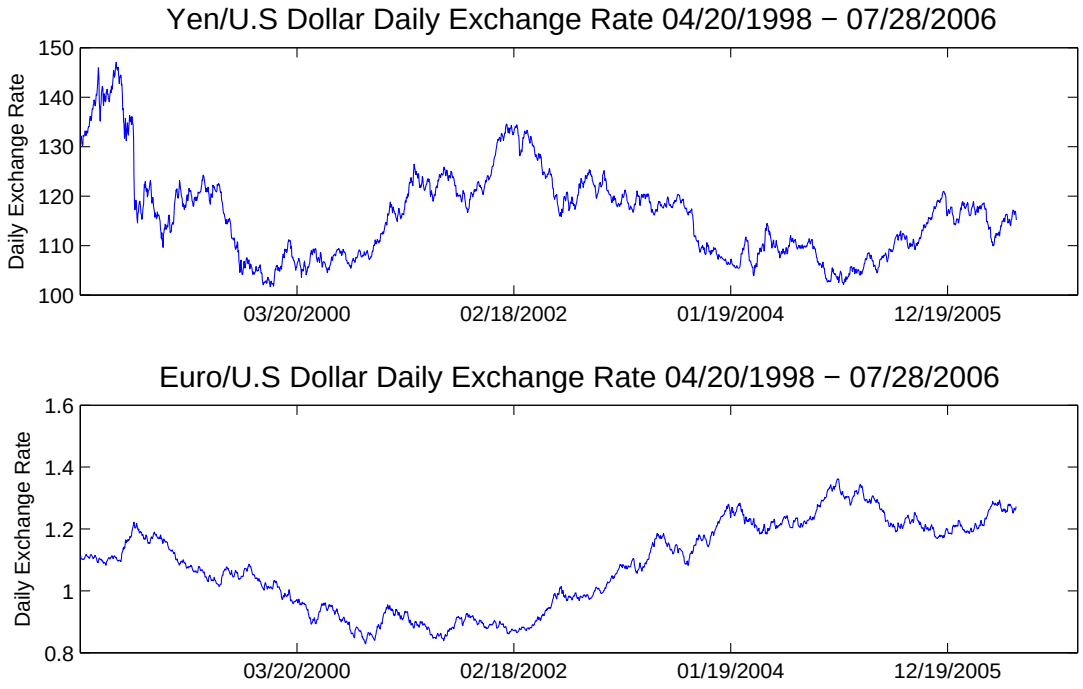

Figure 4.2: The Yen/U.S. Dollar and the Euro/U.S. Dollar daily exchange rates from April 20th, 1998 to July 28th, 2006.

[4]The data are kindly provided by Dr. Tony Wirjanto, University of Waterloo, Canada.

Next, we transform the exchange rate into logarithmic (or continuously compounded) returns using

$$Y_t = 100 * \log\left(\frac{\text{exchange rate in day t}}{\text{exchange rate in day t-1}}\right), \quad t = 2, \ldots, 2159. \tag{4.25}$$

Denote $\{Y_{t,Y}\}$ and $\{Y_{t,E}\}$ respectively as the log return process of the Yen/U.S. Dollar exchange and of the Euro/U.S. Dollar exchange. The new processes $\{Y_{t,Y}\}$ and $\{Y_{t,E}\}$ are shown in Figure 4.3.

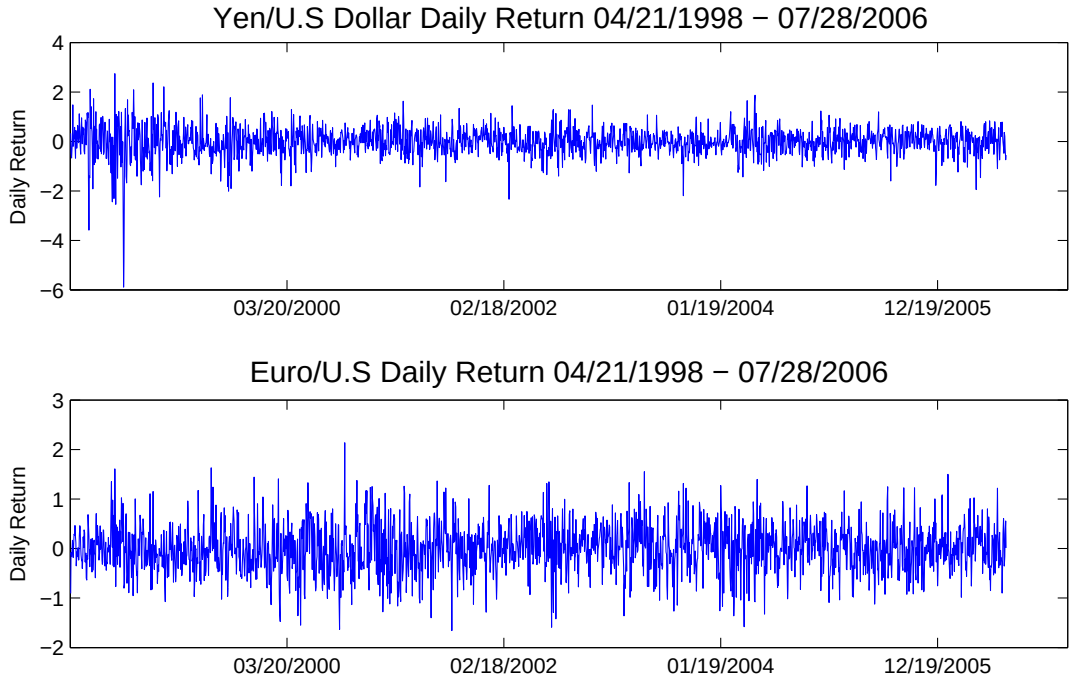

Figure 4.3: The Yen/U.S. Dollar and the Euro/U.S. Dollar daily log returns from April 21st, 1998 to July 28th, 2006.

Both return processes (the upper and the lower graphs in Figure 4.3) display the pattern of

stationarity, clustered volatilities and mean close to zero. The p-values of the Jarque-Bera test (Jarque and Bera, 1980), the Robust Jarque-Bera test (Gel and Gastwirth, 2008) and the SJ test (Gel et al., 2007) are all shown to be less than 0.0001 for $\{Y_{t,Y}\}$ and less than 0.001 for $\{Y_{t,E}\}$, indicating that both the return processes for Yen/USD and Euro/USD are non-Gaussian and hence the new robust permutation and bootstrap tests might be the preferred procedures.

Figure 4.4 exhibits the sample autocorrelations of $\{Y_{t,Y}\}$, $\{Y_{t,E}\}$, $\{Y^2_{t,Y}\}$ and $\{Y^2_{t,E}\}$.

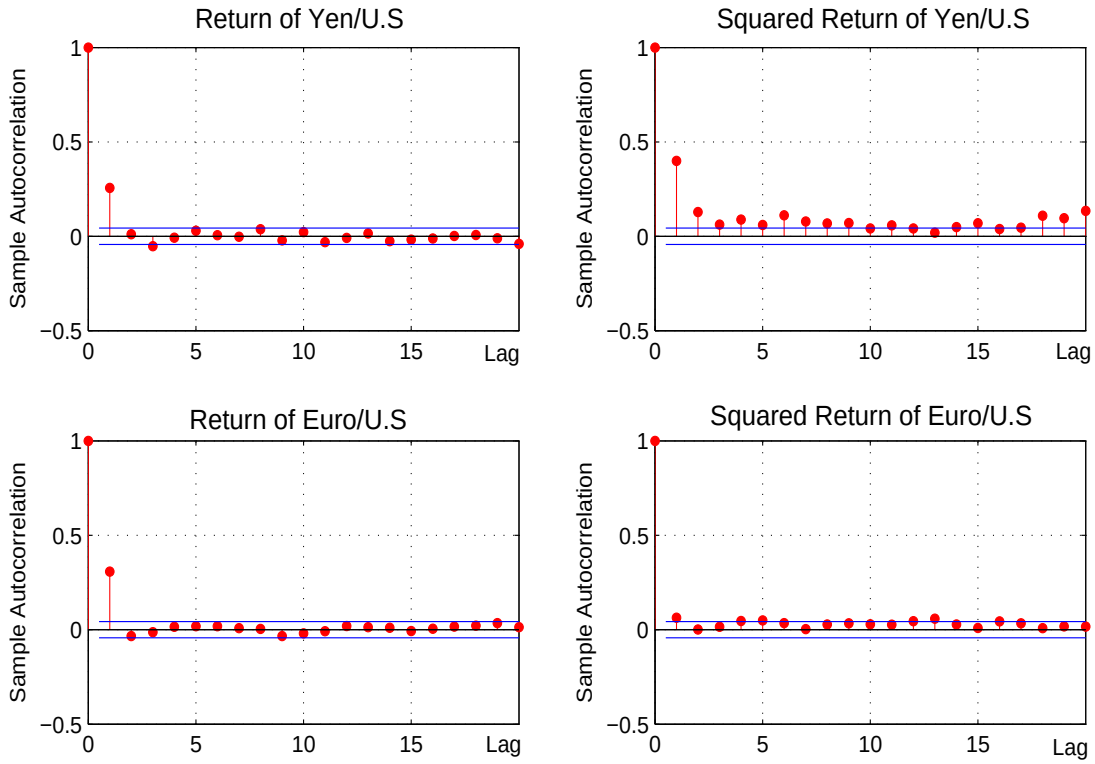

Figure 4.4: Sample autocorrelation plots of the Yen/U.S. Dollar and the Euro/U.S. Dollar daily return processes and squared return processes from April 21st, 1998 to July 28th, 2006.

The return processes $\{Y_{t,Y}\}$ and $\{Y_{t,E}\}$ show significant autocorrelation at lag 1 (the left graphs in Figure 4.4), while both squared return processes $\{Y^2_{t,Y}\}$ and $\{Y^2_{t,E}\}$ have slowly decaying autocorrelations (the right graphs in Figure 4.4). In particular, $\{Y^2_{t,Y}\}$ shows stronger correlations than $\{Y^2_{t,E}\}$. Overall, there is some evidence indicating that $\{Y_{t,Y}\}$ and $\{Y_{t,E}\}$ might have the ARCH/GARCH effect. Hence, we apply the LM tests for a systematic detection.

The left two columns of Table 4.4 show p-values of the PLM, BLM and parametric LM tests for entire samples of $\{Y_{t,Y}\}$ and $\{Y_{t,E}\}$. For the return process $\{Y_{t,Y}\}$ of Yen/U.S. Dollar exchange, the p-values of all three tests are consistently small, i.e. 0 in both parametric and permutation tests and 0.001 for the bootstrap LM test. Hence, there is strong evidence against H_0 that no ARCH effect is present in the daily returns of Yen/USD. As we expected, the p-values of $\{Y_{t,E}\}$ are slightly larger than those of $\{Y_{t,Y}\}$ due to the weaker correlation in squared returns. However, the p-values are still very small and hence we can conclude that an ARCH effect is also present in daily returns of Euro/USD.

In order to demonstrate the performance of the proposed tests in small samples, we take the first 90 observations, roughly corresponding to a three-month duration, from each of $\{Y_{t,Y}\}$ and $\{Y_{t,E}\}$, denoted as $\{Y_{t,Y}\}_{t=1}^{90}$ and $\{Y_{t,E}\}_{t=1}^{90}$ respectively. The right two columns in Table 4.4 present the p-values of the PLM, BLM and parametric LM tests, applied to the sub-samples. Contrary to the results obtained previously for the entire sample from April 20th, 1998 to July 28th, 2006, for the subsample of the first three months of daily returns from the Yen/U.S. Dollar exchange rates $\{Y_{t,Y}\}_{t=1}^{90}$, the parametric LM test yields p-value of 0.095, indicating no evidence for ARCH effect; while both PLM and BLM tests provide statistically significant p-values of 0.041 and 0.037, respectively. Since in small and moderate samples of non-normal processes, according to the simulation studies in Section 3, the new robust permutation and bootstrap LM tests are shown to be more reliable than the parametric LM test, we incline to conclude that there

is an ARCH effect in the first 90 daily returns of Yen/USD, which the parametric LM test fails to detect. On the other hand, for $\{Y_{t,E}\}_{t=1}^{90}$ all p-values are larger than 0.05; hence, there is no evidence of conditional heteroscedasticity in the first 90 daily returns of the Euro/U.S. Dollar exchange rates.

	Entire sample		First 90 observations	
	$\{Y_{t,Y}\}_{t=1}^{2158}$	$\{Y_{t,E}\}_{t=1}^{2158}$	$\{Y_{t,Y}\}_{t=1}^{90}$	$\{Y_{t,E}\}_{t=1}^{90}$
Parametric	0.000	0.003	0.095	0.277
Permutation	0.000	0.002	0.037	0.394
Bootstrap	0.001	0.005	0.041	0.393

Table 4.4: The p-values of the parametric, permutation and bootstrap LM tests for the log return processes of Yen/U.S. Dollar and Euro/U.S. Dollar exchange rates as well as their sub-samples of the first 90 observations. The permutation and the bootstrap LM tests are conducted with 1000 replications.

4.4 Discussion

When the underlying process is heavy-tailed or skewed, which is often the case in practice especially in financial and environmental applications, it is well known that the parametric Lagrange Multiplier (LM) test for conditional heteroscedasticity based on the asymptotic χ^2-distribution is very conservative, i.e. the observed size of the LM test is substantially smaller than the nominal level (Engle et al., 1985; Gregory, 1989; Peguin-Feissolle, 1999; Godfrey et al., 2004; Raunig, 2008). In this chapter, we propose and justify two remedies, namely permutation and bootstrap techniques, to improve performance of the LM test. Our numerical studies indicate that in fi-

nite samples, both new re-sampling based LM tests, especially the permutation LM (PLM) test, effectively and significantly improve the size and power in all the cases considered, even when the data are normal under H_0. Also, we theoretically justify that the PLM test is exact and the bootstrap LM (BLM) test is asymptotically correct. Moreover, although typically the computational time is a practical concern of any re-sampling based methods, especially in a non-linear setting of ARCH/GARCH models, our procedures utilize only linear forms of ARCH/GARCH processes and are therefore computationally efficient and feasible in real-time modeling. For instance, using an Intel Core Duo L2400 1.66GHz processor, the CPU times of conducting PLM and BLM in Matlab for the Yen/U.S. Dollar and the Euro/U.S. Dollar exchange data are 13.12 and 12.23 seconds respectively.

Overall, we conclude that the newly proposed re-sampling based LM tests, particularly PLM, can be viewed as preferred procedures for assessing conditional heteroscedasticity in a variety of applications. Moreover, BLM might be further extended beyond hypothesis testing and applied for sensitivity analysis, model diagnostics as well as assessment of a full predictive distribution density of future returns and volatilities, along the ideas proposed by Pascual et al.(2006) and Chen et al. (2010).

Appendix

Recall that under H_0,

$$Z_t = \alpha_0 + v_t, \text{ where } v_t \sim \text{i.i.d } (0, \tau^2), \tag{4.26}$$

which is equivalent to an auxiliary AR(p) model

$$Z_t - \alpha_0 = \beta_1(Z_{t-1} - \alpha_0) + \ldots + \beta_p(Z_{t-p} - \alpha_0) + v_t, \tag{4.27}$$

where $\beta_1 = \ldots = \beta_p = 0$. Denote $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$, then $\boldsymbol{\beta} = \mathbf{0}$. By equation (4.18), the LS estimate of α_0 is $\bar{z}_n$ and hence, the LS estimate of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{Z} - \bar{\boldsymbol{Z}}). \tag{4.28}$$

Notice that T_n in equation (4.8) can be re-written as

$$T_n = \frac{(\boldsymbol{Z} - \bar{\boldsymbol{Z}})'\boldsymbol{U}(\boldsymbol{U}'\boldsymbol{U})^{-1}(\boldsymbol{U}'\boldsymbol{U})(\boldsymbol{U}'\boldsymbol{U})^{-1}\boldsymbol{U}'(\boldsymbol{Z} - \bar{\boldsymbol{Z}})}{n^{-1}(\boldsymbol{Z} - \bar{\boldsymbol{Z}})'(\boldsymbol{Z} - \bar{\boldsymbol{Z}})}. \tag{4.29}$$

By (4.28), T_n becomes

$$T_n = \frac{\hat{\boldsymbol{\beta}}'(\boldsymbol{U}'\boldsymbol{U})\hat{\boldsymbol{\beta}}}{n^{-1}(\boldsymbol{Z} - \bar{\boldsymbol{Z}})'(\boldsymbol{Z} - \bar{\boldsymbol{Z}})}. \tag{4.30}$$

Proof of Lemma 4.1.1 By Theorem 10.8.2 in Brockwell and Davis (1991),

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(\mathbf{0}, \tau^2[E\boldsymbol{U}'\boldsymbol{U}]^{-1}), \tag{4.31}$$

but $\boldsymbol{\beta} = \mathbf{0}$, hence

$$n^{1/2}\hat{\boldsymbol{\beta}} \xrightarrow{d} N(\mathbf{0}, \tau^2[E\boldsymbol{U}'\boldsymbol{U}]^{-1}). \tag{4.32}$$

Under H_0, $Z_1, \ldots, Z_t$ are i.i.d and $EZ_t = \alpha_0 < \infty$. By the Strong Law of Large Numbers (Kolmogorov), as $n \to \infty$,

$$n^{-1}(\boldsymbol{Z} - \bar{\boldsymbol{Z}})'(\boldsymbol{Z} - \bar{\boldsymbol{Z}}) \to \tau^2 \quad a.s., \tag{4.33}$$

and

$$n^{-1}(\boldsymbol{U}'\boldsymbol{U}) \to E\boldsymbol{U}'\boldsymbol{U} \quad a.s.. \tag{4.34}$$

Therefore, by Slutsky's theorem,

$$n^{1/2}\hat{\boldsymbol{\beta}}'\left(\frac{\boldsymbol{U}'\boldsymbol{U}}{(\boldsymbol{Z} - \bar{\boldsymbol{Z}})'(\boldsymbol{Z} - \bar{\boldsymbol{Z}})}\right)^{1/2} \xrightarrow{d} N(\mathbf{0}, \boldsymbol{I}), \tag{4.35}$$

and the quadratic form in (4.30)

$$\frac{\hat{\boldsymbol{\beta}}'(\boldsymbol{U}'\boldsymbol{U})\hat{\boldsymbol{\beta}}}{n^{-1}(\boldsymbol{Z}-\bar{\boldsymbol{Z}})'(\boldsymbol{Z}-\bar{\boldsymbol{Z}})} \xrightarrow{d} \chi^2_p, \tag{4.36}$$

which completes the proof. □

Proof of Theorem 4.1.3 Similar to equation (4.29), the bootstrap test statistic can be expressed as

$$T^*_n = \frac{(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)'\boldsymbol{U}^*(\boldsymbol{U}^{*'}\boldsymbol{U}^*)^{-1}(\boldsymbol{U}^{*'}\boldsymbol{U}^*)(\boldsymbol{U}^{*'}\boldsymbol{U}^*)^{-1}\boldsymbol{U}^{*'}(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)}{n^{-1}(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)'(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)} \tag{4.37}$$

where $(\boldsymbol{U}^{*'}\boldsymbol{U}^*)^{-1}\boldsymbol{U}^{*'}(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)$ is the bootstrap LS estimate of $\hat{\boldsymbol{\beta}}$, denoted as $\hat{\boldsymbol{\beta}}^*$, then T^*_n becomes

$$T^*_n = \frac{\hat{\boldsymbol{\beta}}^{*'}(\boldsymbol{U}^{*'}\boldsymbol{U}^*)\hat{\boldsymbol{\beta}}^*}{n^{-1}(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)'(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)}. \tag{4.38}$$

By Theorem 4.1 of Kreiss and Franke (1992), as $n \to \infty$,

$$n^{1/2}(\hat{\boldsymbol{\beta}}^*-\hat{\boldsymbol{\beta}}) \xrightarrow{d^*} N(\mathbf{0}, \tau^2[E\boldsymbol{U}'\boldsymbol{U}]^{-1}) \quad \text{in probability}, \tag{4.39}$$

where $\xrightarrow{d^*}$ denotes the convergence in conditional (bootstrap) distribution. Also, $\boldsymbol{\beta}=\mathbf{0}$ under H_0 and hence, by Theorem 10.8.1 of Brockwell and Davis (1991)

$$\hat{\boldsymbol{\beta}} \to \mathbf{0} \quad a.s., \tag{4.40}$$

then we have

$$n^{1/2}\hat{\boldsymbol{\beta}}^* \xrightarrow{d^*} N(\mathbf{0}, \tau^2[E\boldsymbol{U}'\boldsymbol{U}]^{-1}) \quad \text{in probability}. \tag{4.41}$$

In addition, by Theorem 2.1 of Bickel and Freedman (1981), for $\kappa > 0$

$$P^*\left(\left|n^{-1}(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)'(\boldsymbol{Z}^*-\bar{\boldsymbol{Z}}^*)-\tau^2\right| > \kappa\right) \to 0 \quad a.s. \tag{4.42}$$

Also, by Theorem 4.1 of Kreiss and Franke (1992) together with Lemma 8.4 of Bickel and Freedman (1981), for $c > 0$,

$$P^*\big(\big|n^{-1}(\boldsymbol{U}^{*'}\boldsymbol{U}^*) - E\boldsymbol{U}^{'}\boldsymbol{U}\big| > c\big) \to 0 \quad \text{in probability.} \tag{4.43}$$

Therefore, by bootstrap version of Slutsky's theorem,

$$n^{1/2}\hat{\boldsymbol{\beta}}^{*'}\left(\frac{\boldsymbol{U}^{*'}\boldsymbol{U}^*}{(\boldsymbol{Z}^* - \bar{\boldsymbol{Z}}^*)'(\boldsymbol{Z}^* - \bar{\boldsymbol{Z}}^*)}\right)^{1/2} \xrightarrow{d^*} N(\boldsymbol{0}, \boldsymbol{I}) \quad \text{in probability,} \tag{4.44}$$

and the quadratic form in (4.38)

$$\frac{\hat{\boldsymbol{\beta}}^{*'}(\boldsymbol{U}^{*'}\boldsymbol{U}^*)\hat{\boldsymbol{\beta}}^*}{n^{-1}(\boldsymbol{Z}^* - \bar{\boldsymbol{Z}}^*)'(\boldsymbol{Z}^* - \bar{\boldsymbol{Z}}^*)} \xrightarrow{d^*} \chi^2_p \quad \text{in probability.} \tag{4.45}$$

By Lemma 4.1.1, the distribution of T_n and the conditional distribution of T_n^* asymptotically coincide and the result follows. □

Chapter 5

An ANOVA-type Nonparametric Trend Test under GARCH Effect

Tracking and detecting trend play an important role in providing accurate and reliable data analysis and have a variety of applications in financial, environmental, medical and social sciences, e.g., monitoring air/water pollution, tracking global warming, recognizing stock patterns, detecting abnormal heart rhythms, etc. In recent years, trend detection has received considerable attention and a large number of tests have been proposed in the literature (Hipel et al. 1988; Burn, 1994; Burn and Hag Elnur, 2002; Hofmann and Balakrishnan, 2006; Hamed, 2008 and 2009; Kuwabara and Watanabe, 2006; Engle and Rangel, 2008). Overall, the Mann-Kendall test (Mann, 1945; Kendall, 1955) continues to be one of the most common choices for detecting trends, due to its simplicity and nonparametric features. However, it is well-known that Mann-Kendall is designed for assessing monotonic changes in the mean level of the data, while in practice, we often encounter non-monotonic trends particularly in environmental and econometric data where changes in the mean are typically accompanied with irregular seasonal cycles. In

addition, the ARCH/GARCH-type patterns, i.e., conditional heteroscedasticity, usually appear in water quality, streamflow, stock price, foreign exchange rate type time series from environmental and finance studies (Wang, et al., 2005; Wang, 2006; Chen, et al., 2008; Pianosi and Raso, 2008, Anderson, 1998). Our empirical findings show that the power of Mann-Kendall to detect various type of non-monotonic trend under ARCH/GARCH effect is close to zero.

In this chapter, we introduce a new nonparametric trend test in the presence of ARCH/GARCH effect (from here on referred to as NT-GARCH), based on the one-way analysis of variance (ANOVA) F-statistic with large number of groups. NT-GARCH can effectively detect non-monotonic trends under ARCH/GARCH effect, especially when irregular seasonality appears. Our test is an extension of the trend test (NT) suggested by Wang et al. (2008) and Wang and Van Keilegom (2007), where the innovations are assumed either to be heteroscedastic or to be serially correlated. Thus, NT-GARCH inherits all the nonparametric properties of the original NT test, e.g., only a weak smooth assumption is imposed on the functional form of trend and the observed data are not required to be independent and normally distributed. In particular, we suggest to choose the group size by a bootstrap adaptive rule (Bickel and Sakov, 2008). Our simulation studies show that NT-GARCH well-approximates the nominal significance level under the null hypothesis for large samples. For small or moderate samples, we suggest to use a bootstrap procedure to obtain the critical values. Our theoretical findings show that the NT-GARCH statistic is asymptotically normally distributed and the bootstrap NT-GARCH test is consistent. The proposed testing procedure is demonstrated by environmental studies on the water quality of Turkey Lake Watershed and air pollution along Windsor-Quebec City corridor.

5.1 Test statistics and its asymptotic distribution

Let us consider the following nonparametric regression model

$$Y_t = g(x_t) + \epsilon_t, \quad t = 1, \ldots, n, \tag{5.1}$$

where $g(\cdot)$ is an unknown smooth regression function; $x_1 < \ldots < x_n$ is an increasing sequence in $[0, 1]$; the innovations $\{\epsilon_t\}$ follow a GARCH(p, q) process

$$\epsilon_t = \sigma_t e_t, \tag{5.2}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{5.3}$$

where $\{e_t\}$ are i.i.d. random variables with $Ee_t = 0$ and $Ee_t^2 = 1$. We assume that $\{\epsilon_t\}$ is strictly stationary. Note that Wang et al. (2008) and Wang and Van Keilegom (2007) assume respectively that $\{\epsilon_t\}$ is heteroscedastic, where $\epsilon_t = \sigma(x_t)e_t$, and follows an AR($p$) process, where $\epsilon_t = \sum_{i=1}^{p} a_i \epsilon_{t-i} + e_t$. Our objective is to determine whether trend is present in $\{Y_t\}$, i.e., to test the null hypothesis of a constant regression function,

$$H_0 : g(x) = C, \quad \text{for all } x, \tag{5.4}$$

for some unknown constant $C \in \mathbb{R}$.

The NT-GARCH test is motivated by the asymptotic theory for ANOVA with a large number of factor levels (Akritas and Papadatos, 2004; Wang and Van Keilegom, 2007; Wang et al., 2008, and references therein). The idea of NT-GARCH is to consider each time point t as a "factor level" and the symmetric time window W_t around t of length k_n as a "group", where k_n is an odd positive integer, i.e.

$$W_t = \big\{i : |i - t| \leq (k_n - 1)/2\big\}. \tag{5.5}$$

Next, we stack the n groups up to construct an artificial balanced one-way ANOVA with the responses in the t-th group being $\{Y_i, i \in W_t\}$. (Note that the neighbor groups in the ANOVA have common observations; also, for a given sample $\{Y_1, \ldots, Y_n\}$, there are $n-k_n+1$ symmetric windows, that is, $\{W_{(k_n+1)/2}, \ldots, W_{n-(k_n-1)/2}\}$.) Adopting the notations of Wang et al.(2008), we denote V_{ti}, $i = 1, \ldots, k_n$, as the k_n observations in W_t, $0 \leq t \leq n$. Then, the NT-GARCH test statistic is defined as

$$\begin{aligned} T_n &= \sqrt{n/k_n}\,(MST - MSE) && (5.6) \\ &= \sqrt{n/k_n}\left(\frac{k_n}{n-1}\sum_{t=1}^{n}(\bar{V}_{t.} - \bar{V}_{..})^2 - \frac{1}{n(k_n-1)}\sum_{t=1}^{n}\sum_{i=1}^{k_n}(V_{ti} - \bar{V}_{t.})^2\right), && (5.7) \end{aligned}$$

where $\bar{V}_{t.} = k_n^{-1}\sum_{i=1}^{k_n} V_{ti}$ and $\bar{V}_{..} = n^{-1}\sum_{t=1}^{n}\bar{V}_{t.}$, i.e., MST is the treatment sum of squares and MSE is the error sum of squares from the hypothetical one-way ANOVA. Let us denote the vector of all observations in ANOVA as

$$\boldsymbol{V} = (Y_i,\ i \in W_1, \ldots,\ Y_i,\ i \in W_n)'; \tag{5.8}$$

then T_n can be expressed as a quadratic form

$$T_n = \sqrt{n/k_n}(\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V}), \tag{5.9}$$

where

$$\boldsymbol{A} = \frac{nk_n - 1}{n(n-1)k_n(k_n-1)}\bigoplus_{i=1}^{n}\boldsymbol{J}_{k_n} - \frac{1}{n(n-1)k_n}\boldsymbol{J}_{nk_n} - \frac{1}{n(k_n-1)}\boldsymbol{I}_{nk_n} \tag{5.10}$$

where $\boldsymbol{J}_d$ is a $d \times d$ matrix with all elements equal to 1; $\boldsymbol{I}_d$ is a $d \times d$ identity matrix; $\bigoplus$ denotes the Kronecker sum [1].

[1]The Kronecker sum is the matrix sum defined by

$$\boldsymbol{A}\bigoplus\boldsymbol{B} = \boldsymbol{A}\bigotimes\boldsymbol{I}_b + \boldsymbol{I}_a\bigotimes\boldsymbol{B}, \tag{5.11}$$

The rest of this section contributes to the study of asymptotic properties of T_n under H_0 in two frameworks: (1) $n \to \infty$ and $k_n \to \infty$ at appropriate rate; (2) $n \to \infty$ and k_n is fixed. We assume without loss of generality that $C = 0$ in the following discussion. Note that, under H_0, $\{Y_t\} = \epsilon_t$ and follows a GARCH(p, q) model 5.2.

Let us first state a lemma which allows us to consider a much simpler statistic that is asymptotically equivalent to T_n. Note that this result generalizes Lemma 3.1 of Wang et al. (2008) which assumes that $\{Y_t\}$ is an independent process.

Lemma 5.1.1 *Assume that e_t is symmetrically distributed and $E(e_t^4) < \infty$. Under H_0,*

$$\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A} - \boldsymbol{A}_d)\boldsymbol{V} \xrightarrow{P} 0 \tag{5.12}$$

as $n \to \infty$ and $k_n/n \to 0$, where $\boldsymbol{A}_d$ is the block diagonal matrix

$$\boldsymbol{A}_d = \text{diag}\{\boldsymbol{B}_1, \ldots, \boldsymbol{B}_n\}, \text{ with } \boldsymbol{B}_i = \frac{1}{n(k_n - 1)}[\boldsymbol{J}_k - \boldsymbol{I}_k]. \tag{5.13}$$

The proof of Lemma 5.1.1 is given in the Appendix at the end of this chapter.

The next theorem establishes the asymptotic variance of the statistic $\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}$ under H_0. Let us denote

$$\gamma(h) \quad = \quad E(Y_t^2 Y_{t+h}^2)$$

for $h \in \mathbb{Z}$. Note that, because the process $\{Y_t, t \in \mathbb{Z}\}$ is strictly stationary, $E(Y_t^2 Y_{t+h}^2)$ does not depend on t but only on the lag h.

---

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are respectively $a \times a$ and $b \times b$ square matrices and $\bigotimes$ denotes the Kronecker product (Horn and Johnson, 1994).

Theorem 5.1.2 *Suppose that e_t is symmetrically distributed and $E(e_t^4) < \infty$. Under H_0, as $n \to \infty$,*

$$Var\big(\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}\big) \to \tau^2$$

where

(1) if $k_n = k$ is fixed, then

$$\tau^2 = \frac{4}{k(k-1)^2}\sum_{h=1}^{k-1}(k-h)^2\gamma(h). \tag{5.14}$$

(2) if $k_n \to \infty$ such that $2k_n \le n$ and $k_n/n \to 0$, then

$$\tau^2 = \frac{4}{3}\alpha_0^2\bigg(1 - \sum_{i=1}^{p}\alpha_i - \sum_{j=1}^{q}\beta_j\bigg)^{-2}. \tag{5.15}$$

The detailed proof of Theorem 5.1.2 is provided in Appendix.

Remark 1. Under the assumption k_n is fixed, a closed form of $\gamma(h)$ can be derived for some special cases of GARCH(p, q) processes. For example,

- If $\{Y_t, t \in \mathbb{Z}\}$ follows an ARCH(1) process, then

$$\gamma(h) = \alpha_1^h\bigg\{\frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)} - \bigg(\frac{\alpha_0}{1-\alpha_1}\bigg)^2\bigg\} + \bigg(\frac{\alpha_0}{1-\alpha_1}\bigg)^2 \tag{5.16}$$

- If $\{Y_t, t \in \mathbb{Z}\}$ follows a GARCH(1, 1) process, then

$$\gamma(h) = \frac{\alpha_1(1-\alpha_1\beta_1-\beta_1^2)}{1-(\alpha_1+\beta_1)^2+\alpha_1^2}(\alpha_1+\beta_2)^{h-1}\gamma(0) + \bigg(\frac{\alpha_0}{1-\alpha_1-\beta_1}\bigg)^2, \tag{5.17}$$

where

$$\gamma(0) = \frac{2\alpha_0^2(1+\alpha_1+\beta_1)}{(1-\alpha_1-\beta_1)(1-2\alpha_1^2-\beta_1^2-2\alpha_1\beta_1)}\bigg(1 + \frac{\alpha_1^2}{1-(\alpha_1+\beta_1)^2}\bigg). \tag{5.18}$$

However, $\gamma(h)$ has a very complicated form (if a closed expression exists) for a GARCH(p, q) process when p, $q > 1$. Therefore, we suggest, in general, to apply the asymptotic variance τ^2 where we allow $k_n \to \infty$ or to utilize the bootstrap approach which we introduce in the next section.

In the following theorem, we establish asymptotic normality of the test statistic T_n when k_n is fixed.

Theorem 5.1.3 *Suppose that e_t is symmetrically distributed and $E|e_t|^{4+\delta} < \infty$ for some $\delta > 0$. If $k_n = k$ is fixed, under H_0,*

$$T_n \to N(0,\ \tau^2),$$

as $n \to \infty$, where τ^2 is given in Theorem 5.1.2 part (1).

The proof of Theorem 5.1.3 is based on the asymptotic equivalence of $\boldsymbol{V}'\boldsymbol{A}\boldsymbol{V}$ and $\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}$ result of Lemma 5.1.1. The detailed proof is given in Appendix.

In the case $k_n \to \infty$ as $n \to \infty$, we state the following conjecture.

Conjecture 2 *Suppose that e_t is symmetrically distributed and $E|e_t|^{4+\delta} < \infty$ for some $\delta > 0$. Under H_0, as $n \to \infty$ and $k_n \to \infty$ such that $2k_n \leq n$ and $k_n/n \to 0$,*

$$T_n \to N(0,\ \tau^2),$$

where τ^2 is given in Theorem 5.1.2 part (2).

Our key idea to prove Conjecture 2 is to decompose $\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}$ into a sum of "blocks" U_{n1}, V_{n1}, U_{n2}, V_{n2}, ... such that the V_{ni} blocks are asymptotically negligible, and yet long enough to avoid

any overlap among the terms in the U_{ni} blocks, i.e.,

$$\sqrt{\frac{n}{k_n}}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V} \;=\; \frac{1}{\sqrt{nk_n}}\left(\sum_{i=1}^{r_n} U_{ni} + \sum_{i=1}^{r_n} V_{ni}\right),$$

where $r_n = O\big(n^{1/3}k_n^{-1/3}\big)$. To establish the result, it is sufficient to show that $(nk_n)^{-1/2}\sum_{i=1}^{r_n} U_{ni}$ converges to $N(0,\ \tau^2)$. However, we encountered major technical difficulty verifying the uniform integrability of $(nk_n)^{-1/2}U_{ni}$, which is required by the Lindeberg's central limit theorem for strongly mixing triangular sequences. One possible remedy is to restrict e_t to be a sub-Gaussian random variable. After numerous unsuccessful attempts, we decide to leave it as a conjecture. Figures 5.1 and 5.2 provide empirical evidence to support Conjecture 2. Here we consider an ARCH(1) process with parameters $(\alpha_0, \alpha_1)' = (0.1, 0.3)'$, whose asymptotic variance using equation (5.15) is 0.0272. The window length k_n is chosen approximately to be $n^{1/2}$ in Figures 5.1 and $n^{3/5}$ in Figures 5.2, both increases with sample size n. As the plots show, the distribution of T_n becomes closer to normal as n increases. Also, the simulated variances of T_n are 0.0282, 0.0266, 0.0274 respectively for sample sizes 1000, 5000, 10000 in Figures 5.1 and 0.0247, 0.0252, 0.0264 in Figures 5.2 based on 5000 Monte Carlo (MC) simulations, which clearly indicates the pattern of converging to the asymptotic variance.

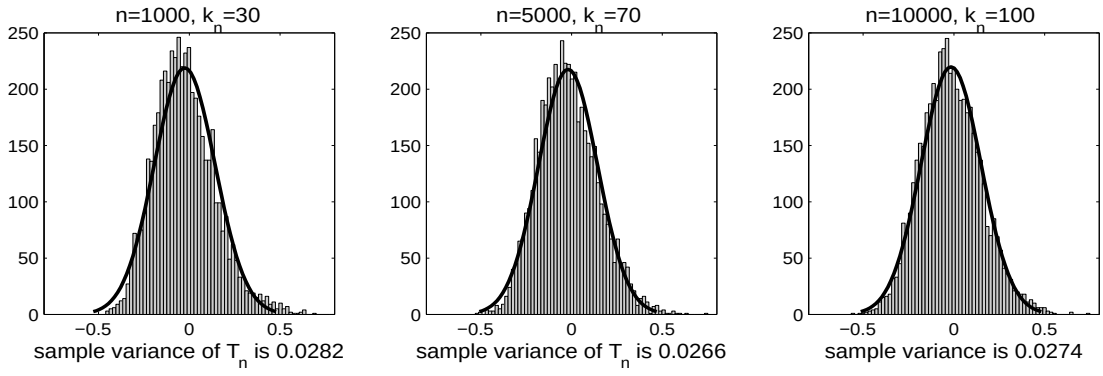

Figure 5.1: The distributions of T_n of the ARCH(1) process with parameters $(\alpha_0, \alpha_1)' = (0.1, 0.3)'$ for sample sizes n to be 1000, 5000 and 10000 respectively. Each window length k_n is chosen approximately to be $\sqrt{n}$. The number of Monte Carlo simulations is 5000.

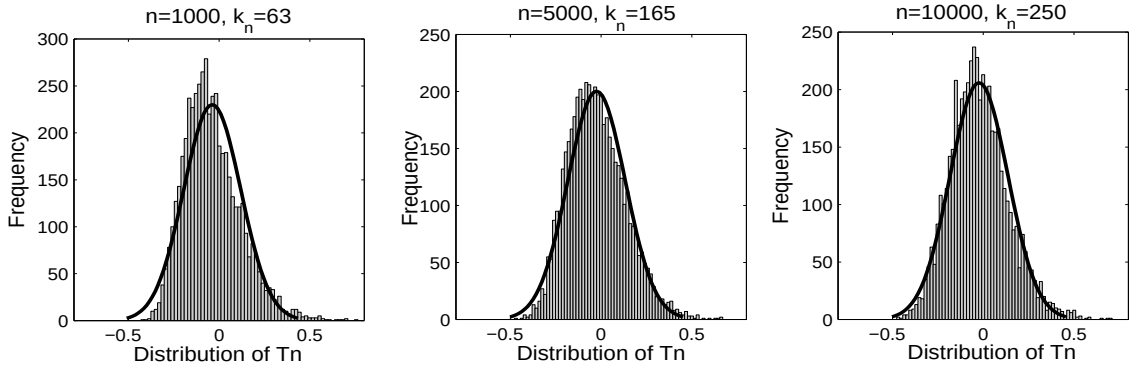

Figure 5.2: The distributions of T_n of the ARCH(1) process in Figure 5.1. Each window length k_n is chosen approximately to be $n^{3/5}$. The number of Monte Carlo simulations is 5000.

Remark 2. The null hypothesis can be generalized to assuming that $g(x)$ belongs to some parametric family of functions $S_\Theta = \{g(., \theta), \theta \in \Theta\}$ where Θ is a subset of a Euclidean space

and $g(.,\theta)$ is defined on $\mathbb{R}$. Suppose $S_{\Theta_0} \in S_\Theta$. The hypothesis test is

$$H_0 : \quad g(x,\theta) \in S_{\Theta_0}, \tag{5.19}$$

$$H_1 : \quad g(x,\theta) \in S_\Theta \backslash S_{\Theta_0}. \tag{5.20}$$

Consequently, NT-GARCH can be used as a goodness-of-fit test, e.g., distinguishing a linear trend from a quadratic one. In addition, our empirical results show that NT-GARCH is useful for detecting so called "stochastic trend", e.g.,

$$Y_t = \mu + Y_{t-1} + \epsilon_t, \tag{5.21}$$

where $\mu \in \mathbb{R}$ and ϵ_t follows a GARCH process.

5.2 The bootstrap NT-GARCH test and asymptotic results

Note that when the window $k_n = k$ is fixed, the asymptotic variance of the NT-GARCH statistic depends on k. However, there has been no theoretically solid method to choose k so far (an empirical bootstrap algorithm for selecting k will be discussed later in this section). Moreover, our simulations indicate that the asymptotic version of the NT-GARCH test tends to be liberal in small and moderate samples. Therefore, we suggest to utilize the bootstrapped version of the NT-GARCH test, obtained by the following residual-based bootstrap algorithm (Maercker and Moser, 2009).

Step 1. Estimate the GARCH coefficients $(\hat{\alpha}_0, \hat{\alpha}_1, \ldots, \hat{\alpha}_p, \hat{\beta}_1, \ldots, \hat{\beta}_q)$ using the Least Square (LS) method based on the ARMA(p,q) representation of $\{Y_t^2\}$

$$Y_t^2 = \alpha_0 + \sum_{i=1}^{m} (\alpha_i + \beta_i) Y_{t-i}^2 + v_t - \sum_{j=1}^{q} \beta_j v_{t-j}, \tag{5.22}$$

where $m = \max(p, q)$, $\alpha_i = 0$ for $i > p$ and $\beta_i = 0$ for $i > q$; $v_t = \sigma_t^2(e_t^2 - 1)$ is white noise.

Step 2. Compute the estimated volatility process $\{\hat{\sigma}_t^2,\ m+1 \le t \le n\}$ by

$$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i Y_{t-i}^2 + \sum_{j=1}^{q} \hat{\beta}_j \hat{\sigma}_{t-j}^2, \tag{5.23}$$

Step 3. Estimate the residuals $\{\hat{e}_t,\ m+1 \le t \le n\}$ by

$$\hat{e}_t = Y_t / \hat{\sigma}_t. \tag{5.24}$$

In addition, let

$$\hat{G}_{e,n}^{+}(x) = \frac{1}{n-m} \sum_{t=m+1}^{n} \hat{e}_t, \quad \hat{G}_{e,n}^{-}(x) = \frac{1}{n-m} \sum_{t=m+1}^{n} (-\hat{e}_t), \tag{5.25}$$

then the empirical distribution of $\hat{e}_t$ is

$$\hat{G}_{e,n}(x) = \left\{\hat{G}_{e,n}^{+}(x) + \hat{G}_{e,n}^{-}(x)\right\}/2. \tag{5.26}$$

Note that by such construction, $\hat{G}_{e,n}(x)$ is symmetric.

Step 4. Smooth $\hat{G}_{e,n}(x)$ by convolution

$$\tilde{G}_{e,h}(x) = \hat{G}_{e,n}(x) * \Phi(x/h), \tag{5.27}$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)dt$ is the probability function $N(0,1)$ and $h = n^{-1/5}$ is the rate of smoothing.

Step 5. Sample with replacement from $\tilde{G}_{e,h}\left(\hat{\sigma}_{\tilde{G}} x + \hat{\mu}_{\tilde{G}}\right)$ to obtain the bootstrap innovations $\{e_t^*, 1 \le t \le n\}$, where

$$\hat{\mu}_{\tilde{G}} = \frac{1}{n-m} \sum_{t=m+1}^{n} \hat{e}_t, \ \ \hat{\sigma}_{\tilde{G}} = \frac{1}{n-m} \sum_{t=m+1}^{n} (\hat{e}_t - \hat{\mu}_{\tilde{G}})^2 + h^2 \tag{5.28}$$

are the mean and variance of $\tilde{G}_{e,n}$, respectively.

Remark 3. Consequently, the probability density of e_t^* is given by

$$\tilde{g}(x) = \frac{\hat{\sigma}_{\tilde{G}}}{nh} \sum_{t=1}^{n} \phi\left(\frac{\hat{\sigma}_{\tilde{G}} x + \hat{\mu}_{\tilde{G}} - \hat{e}_t}{h} \right), \tag{5.29}$$

where $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ is the probability density of $N(0,1)$. Hence, $\tilde{g}$ is actually a standardized kernel estimate of the underlying probability density of e_t with kernel ϕ and bandwidth h.

Step 6. Construct a bootstrap sample $\{Y_t^*, 1 \leq t \leq n\}$ recursively by

$$Y_t^* = \sigma_t^* e_t^*, \tag{5.30}$$

$$\hat{\sigma}_t^{2*} = \hat{\alpha}_0 + \sum_{i=1}^{p} \hat{\alpha}_i Y_{t-i}^{2*} + \sum_{j=1}^{q} \hat{\beta}_j \sigma_{t-j}^{2*}, \tag{5.31}$$

and set $\sigma_k^{2*} = \hat{\alpha}_0^{-1}\Big(1 - (\sum_{i=1}^{p} \hat{\alpha}_i + \sum_{j=1}^{q} \hat{\beta}_j)\Big)$ for $k \leq 0$. In practice we generate $\{Y_t^*\}$ of size $n + 150$ and then discard the first 150 generated values in order to minimize the effect of the initial values. The effect of initial values is negligible asymptotically (Kreiss and Franke, 1992).

Step 7. Denote V_{ti}^*, $i = 1, \ldots, k_n$, as the k_n bootstrap observations in W_t, $0 \leq t \leq n$. Calculate the bootstrap NT-GARCH test statistic by

$$T_{n,b}^* = \sqrt{n/k_n}\Big\{ \frac{k_n}{n-1} \sum_{t=1}^{n} (\bar{V}_{t.}^* - \bar{V}_{..}^*)^2 - \frac{1}{n(k_n-1)} \sum_{t=1}^{n} \sum_{i=1}^{k_n} (V_{ti}^* - \bar{V}_{t.}^*)^2 \Big\}, \tag{5.32}$$

where $\bar{V}_{t.}^* = k_n^{-1} \sum_{i=1}^{k_n} V_{ti}^*$ and $\bar{V}_{..}^* = n^{-1} \sum_{t=1}^{n} \bar{V}_{t.}^*$.

Step 8. Repeat **Steps 4–6** B times to obtain $\{T^*_{n,1}, \ldots, T^*_{n,b}, \ldots, T^*_{n,B}\}$ where B is the number of bootstrap replications and b denotes the b^{th} iteration. Under H_0, the unknown distribution of T_n is approximated by

$$\hat{F}^*_{T_n}(x) = \sum_{b=1}^{B} \mathbf{1}_{\{T^*_{n,b} < x\}}, \tag{5.33}$$

and the unknown variance of T_n is approximated by

$$\hat{\tau}^{*2} = \frac{1}{B-1} \sum_{b=1}^{B} \Big(T^*_{n,b} - \hat{\mu}_{\hat{F}^*} \Big)^2, \tag{5.34}$$

where $\hat{\mu}_{\hat{F}^*} = (B-1)^{-1} \sum_{b=1}^{B} T^*_{n,b}$.

The purpose of applying the residual-bootstrap method is multi-fold. Based on the bootstrap construction of $\hat{F}^*_{T_n}(x)$, we can systematically select the window length k_n using the following adaptive rule (Bickel and Sakov, 2008):

Step a. Let $k_{n,j}$ be a sequence of odd numbers of the form

$$k_{n,j} = \begin{cases} \lfloor \frac{j}{100} n \rfloor, & \text{if } \lfloor \frac{j}{100} n \rfloor \text{is odd,} \\ \\ \lfloor \frac{j}{100} n \rfloor + 1, & \text{if } \lfloor \frac{j}{100} n \rfloor \text{is even,} \end{cases}$$

for $j = 1, \ldots, 100$. Calculate the corresponding sequence of empirical distributions $\hat{F}^*_{T_n, k_{n,j}}$ using the window length $k_{n,j}$.

Step b. Let ρ be some metric consistent with convergence in law. Choose k_n to be

$$k_n = \underset{k_{n,j}}{\operatorname{argmin}}\, \rho\big(\hat{F}^*_{T_n,k_{n,j}}, \hat{F}^*_{T_n,k_{n,j+1}}\big), \tag{5.35}$$

for $j = 1, \ldots, n$. Consequently, the "optimal" distribution estimator of F_{T_n} is $\hat{F}^*_{T_n,k_n}$.

As the result of applying the bootstrap adaptive rule (BAR), we obtain not only the "optimal" k_n but also the corresponding $\hat{F}^*_{T_n,k_n}$. The bootstrap critical value of T_n at significance level α is obtained by taking the upper α-quantile of $\hat{F}^*_{T_n,k_n}$, denoted by $Q^*_{T_n,k_n}(1-\alpha)$, and the null hypothesis of no trend is rejected if $T_n \geq Q^*_{T_n,k_n}(1-\alpha)$.

As our simulation studies indicate, BAR can effectively select an "optimal" k_n which leads to a test size very close to the nominal significance level. In particular, the utilization of the bootstrap critical value corresponding to such "optimal" k_n significantly improves the size of the NT-GARCH test in finite samples. Note that if k_n is assumed to be fixed, then we can omit the adaptive rule procedure for selecting k_n whenever n increases, but directly obtain the bootstrap critical value.

Remark 4. The adaptive rule is initially proposed by Bickel and Sakov (2008) for m out of n bootstrap. We adopt this algorithm into our window length k_n selection. In practice, we let ρ to be the Euclidean norm of the distance between adjacent empirical distributions, i.e.

$$k_n = \underset{k_{n,j}}{\operatorname{argmin}} \left\| \hat{F}^*_{T_n,k_{n,j}} - \hat{F}^*_{T_n,k_{n,j+1}} \right\|. \tag{5.36}$$

In the case of k_n changing with n, our numerical results suggest that k_n should not exceed approximately $10\%n$ in order to maintain the nominal significance level. For example, Figure 5.3 shows sizes of the NT-GARCH test with respect to various n's from 50 to 400 of an ARCH (1) process (left panel) and a GARCH (1,1) process (right panel). The five size curves in each plot correspond to $k_n = 5, 9, 19, 29$ and 39. Given a k_n, both plots indicate that size of the NT-GARCH test monotonically increases with n but at a decreasing rate. When n is about 10 times of k_n, size is close to the nominal significance level $\alpha = 0.05$ or lower, e.g, when $k_n = 9$ and $n = 100$ of the ARCH (1) process, the test size is 0.0532; when $k_n = 19$ and $n = 200$ of the

GARCH (1,1) process, the test size is 0.045. Therefore, 10% of the sample size n may be taken as an upper bound of k_n. In practice, we select k_n within the range $[3,\ 0.15n]$ by BAR.

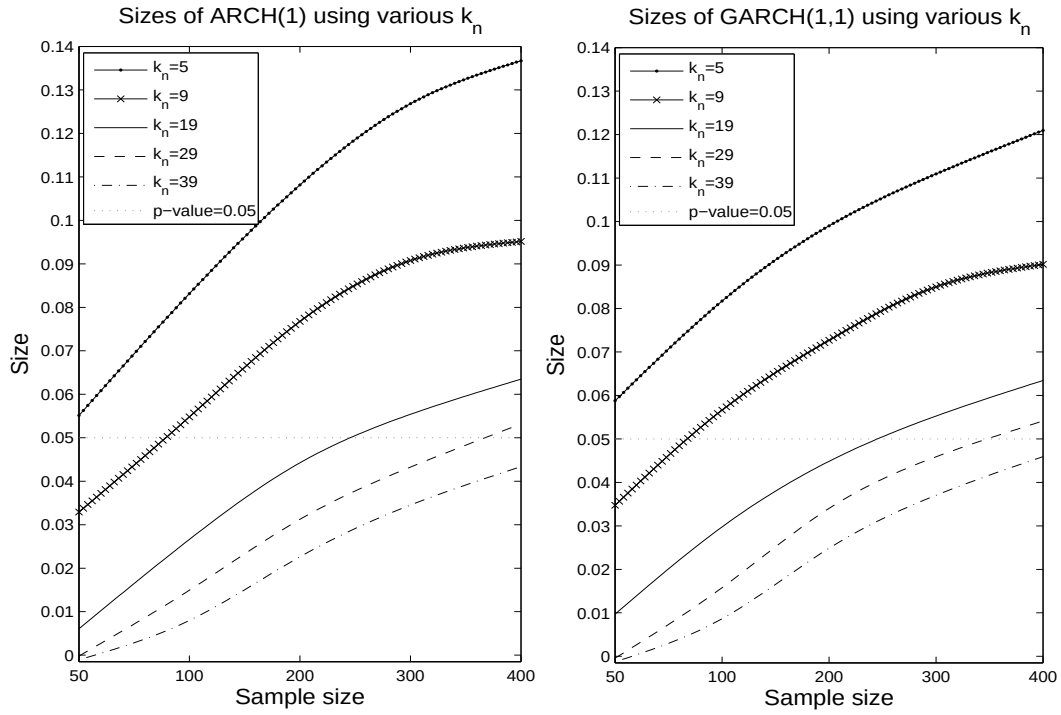

Figure 5.3: Size of the NT-GARCH test for various sample sizes using $k_n = 5,\ 9,\ 19,\ 29$ and 39, respectively, of a ARCH (1) process (left panel) and an GARCH (1,1) process (right panel). Number of Monte Carlo simulations is 5000. The grey dashed line represents the nominal significance level $\alpha = 0.05$.

Next result shows that the bootstrap NT-GARCH test is asymptotically consistent, i.e., it gives the asymptotically correct α level under H_0.

Theorem 5.2.1 *Assume e_t is symmetrically distributed and $E|\epsilon_t|^{4+\delta} < \infty$ for some $\delta > 0$.*

Under H_0, as $n \to \infty$ and $k_n = k$ is fixed

$$\sup_t \left| P^*(\sqrt{n/k_n} T_n^* \leq t) - P(\sqrt{n/k_n} T_n \leq t) \right| = o_p(1), \tag{5.37}$$

where P^* denotes the conditional probability $P(\cdot|Y_1, \ldots, Y_n)$.

The detailed proof of a special case of GARCH(p, q) when $p = q = 1$ is given in Appendix. The proof for general GARCH(p, q) processes can be achieved by using similar arguments.

5.3 Numerical Studies

In this section we present simulation studies on size and power of the proposed NT-GARCH test, based on $\{\epsilon_t\}$ following the ARCH/GARCH models:

- Model 1: ARCH(1)

$$\begin{aligned} \epsilon_t &= \sigma_t e_t, \\ \sigma_t^2 &= 0.1 + 0.4\epsilon_{t-1}^2, \end{aligned}$$

- Model 2: GARCH(1,1)

$$\begin{aligned} \epsilon_t &= \sigma_t e_t, \\ \sigma_t^2 &= 0.15 + 0.1\epsilon_{t-1}^2 + 0.6\sigma_{t-1}^2, \end{aligned}$$

where e_t are independent and identically distributed (i.i.d.) random variables with a continuous distribution function F_e. In order to assess how the newly proposed procedure performs in various distributions, we consider F_e to be a wide range of heavy tailed and skewed distributions, including uniform, exponential, lognormal, laplace, χ^2- and t-distributions as well as Normal

Inverse Gaussian (NIG) distributions with different combinations of parameters and a Tukey Contaminated Normal distribution (CN). Note that all distributions considered are centered with mean zero. All the results are based on 5000 MC iterations and 999 bootstrap replication of each sample, under a nominal level of $\alpha = 0.05$.

Figure 5.4 shows the selections of k_n using BAR for processes following Model 1 with $e_t \sim N(0,1)$ and Model 2 with $e_t \sim logistic$. According to the argument given in the previous section, we restrict the selection range to be $[3,\ 0.15n]$ to optimize computation time, e.g., when $n = 50$, k_n is chosen from $\{3, 5, 7\}$ and when $n = 100$, k_n is chosen from $\{3, 5, 7, 9, 11, 13, 15\}$. Among 5000 MC iterations, the choice of k_n mostly occurs at 5 for $n = 50$ and at 9 for $n = 100$, observed from the histograms in Figure 5.4 in both ARCH(1) (upper plots) and GARCH(1,1) (lower plots) cases; for larger samples, k_n tends to be smaller than $10\%n$, e.g, the average of k_n is 15 for $n = 200$ and 33 for $n = 500$.

5.3.1 Size of the NT-GARCH test

First of all, we investigate the size of the NT-GARCH test when $k_n = k$ is fixed as sample size n increases, for both parametric (PNT-GARCH) and bootstrap (BNT-GARCH) procedures (see Table 5.1 for the size of $\{Y_t\}$ following Model 1 under H_0). Here we select k to be 5 or 9. For all of the symmetric distributions, the sizes of both PNT-GARCH and BNT-GARCH steadily approach the nominal significance level 0.05 as n increases, regardless of the value of k. In small and moderate samples, BNT-GARCH significantly outperforms the parametric one, e.g, for 100 observations from a t_5-distribution, the size of BNT-GARCH is 0.033 comparing to 0.028 yielded by PNT-GARCH ($k = 5$); for 300 observations from a uniform distribution, BNT-GARCH provides size of 0.0548 while PNT-GARCH has size of 0.026 ($k = 9$). Note that

we do not have a theoretical justification on asymptotic normality of the NT-GARCH statistic when $\{e_t\}$ are asymmetrically distributed. Nevertheless, we also investigate performance of parametric and bootstrap versions of the NT-GARCH test, in application to skewed distributions. As our simulations indicate (see Table 5.1), for exponential distribution, the estimated Type I error of the PNT-GARCH monotonically increases with n. (We observe similar patterns for

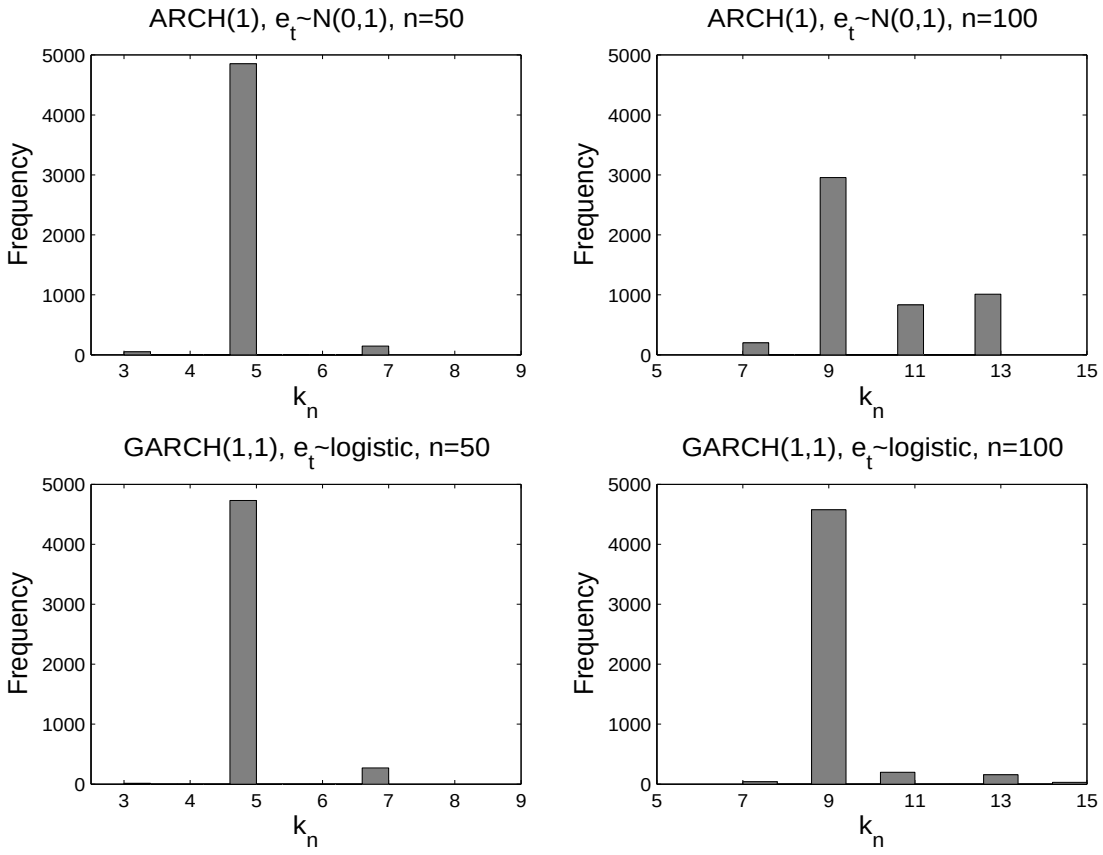

Figure 5.4: The selection of k_n using the bootstrap procedure for Model 1 with $e_t \sim N(0,1)$ (upper panels) and for Model 2 with $e_t \sim logistic$ (lower panels).

other skewed distributions, e.g., lognormal (the results are not included here for brevity.). In contrast, sizes of BNT-GARCH in the exponential case are close to 0.05 for all sample sizes. Therefore in practice, the test using bootstrap well-approximates the nominal significance level even when the data are skewed.

					Sample Size		
Distribution	Skewness	Kurtosis	k_n	Method	100	300	1000
$N(0,1)$	0	3	5	PNT-GARCH	0.029	0.0403	0.046
				BNT-GARCH	0.0398	0.046	0.0472
			9	PNT-GARCH	0.021	0.0362	0.0466
				BNT-GARCH	0.0354	0.042	0.0508
t_5	0	6	5	PNT-GARCH	0.028	0.031	0.0578
				BNT-GARCH	0.033	0.0352	0.0478
			9	PNT-GARCH	0.0292	0.0322	0.0498
				BNT-GARCH	0.0318	0.0326	0.049
$Unif$	0	1.8	5	PNT-GARCH	0.0209	0.028	0.032
				BNT-GARCH	0.0458	0.0542	0.05
			9	PNT-GARCH	0.0215	0.026	0.0274
				BNT-GARCH	0.045	0.0548	0.0474
$Exp(1)$	2	9	5	PNT-GARCH	0.029	0.0841	0.1122
				BNT-GARCH	0.035	0.0506	0.0448
			9	PNT-GARCH	0.0146	0.0573	0.09
				BNT-GARCH	0.031	0.0446	0.0422

Table 5.1: Size of the NT-GARCH test for the processes following Model 1 with e_t respectively being $N(0,1)$, t_5, $Unif$ and $Exp(1)$ random variables at $\alpha = 5\%$ using $MC = 5000$ and $B = 999$. The window length k is fixed to be 5 or 9.

Remark 5. We also conduct a simulation study on size of the original NT (Wang and Van Keilegom, 2007) test, when $\{Y_t\}$ follows an ARCH(1) process in Model 1 under H_0. NT se-

lects window length k to be 5, 7 or 9 (the case $k = 7$ shows similar sizes and hence omitted here). Table 5.2 shows the sizes of both parametric NT (PNT) and bootstrap NT (BNT) test, at significance level of 0.05. Clearly, PNT and BNT severely over-reject H_0 in all of $N(0, 1)$, t_5, exponential and uniform distributions, even for large samples, e.g., for 500 observations from a t_5 distribution, the size of PNT is 0.2982 and that of BNT is 0.1914 ($k = 5$). In addition, the sizes of both PNT and BNT monotonically increase with n. Therefore, it is crucial to take into account the ARCH effect for the trend detection using such ANOVA-type nonparametric tests. Otherwise, it could lead to unreliable or even erroneous conclusions.

					Sample Size	
Distribution	Skewness	Kurtosis	k_n	Method	100	500
$N(0,1)$	0	3	5	PNT	0.1326	0.1592
				BNT	0.0850	0.1052
			9	PNT	0.1012	0.1318
				BNT	0.0622	0.0858
t_5	0	6	5	PNT	0.2386	0.2982
				BNT	0.1144	0.1914
			9	PNT	0.1858	0.2668
				BNT	0.0960	0.1532
$Unif$	0	1.8	5	PNT	0.1196	0.1388
				BNT	0.0724	0.0730
			9	PNT	0.0958	0.1060
				BNT	0.0600	0.0696
$Exp(1)$	2	9	5	PNT	0.1120	0.1532
				BNT	0.0848	0.1005
			9	PNT	0.0828	0.1328
				BNT	0.0776	0.0900

Table 5.2: Size of the NT test for the processes following Model 1 with e_t respectively being $N(0,1)$, t_5, $Unif$ and $Exp(1)$ random variables at $\alpha = 5\%$ using $MC = 5000$ and $B = 999$. The window length k_n is fixed to be 5 or 9.

Next, we discuss the size of NT-GARCH when $k_n \to \infty$ as $n \to \infty$. Table 5.3 and 5.4 show the simulation results on size for symmetric distributions while Table 5.5 and 5.6 for skewed

distributions. In particular, k_n is chosen by BAR for both parametric and bootstrap tests (which implies k_n increases with n). Note that for symmetric distributions, the sizes of BNT-GARCH are generally close to those of PNT-GARCH, so we only include the results of BNT-GARCH in Table 5.3 and 5.4. Overall, the sizes of BNT-GARCH improve noticeably for all considered distributions and sample sizes, compared to the results with fixed k_n (Table 5.1). Notice that even for 50 observations, sizes are extremely close to the nominal significance level 0.05, e.g., the size of BNT-GARCH is equal to 0.0506 when $\{Y_t\}$ follows Model 1 with t_5-distributed $\{e_t\}$ and is 0.065 when $\{Y_t\}$ follows Model 2 with uniformly distributed $\{e_t\}$. The results in Table 5.3 and 5.4 clearly show a pattern that the estimated Type I error converge to 0.05 as n increases.

		Sample Size				
Distribution	Kurtosis	50	100	150	200	500
$N(0,1)$	3	0.0524	0.0532	0.0488	0.0510	0.0502
$CN_{0.25,0.4}$	6.4	0.0504	0.0488	0.0503	0.0500	0.0498
t_5	6	0.0504	0.0512	0.0490	0.0482	0.0488
$Laplace$	6	0.0522	0.0482	0.0466	0.0456	0.0484
$NIG_{0.4,0,1,0.6}$	12.5	0.0544	0.0456	0.0432	0.0428	0.0460
$Unif$	1.80	0.0524	0.0540	0.0512	0.0500	0.0496
$Logistic$	1.20	0.0520	0.0482	0.0462	0.0440	0.0496

Table 5.3: Size of the NT-GARCH test for the processes following Model 1 with various heavy-tailed distributions at $\alpha = 5\%$ using $MC = 5000$ and $B = 999$. The window length k_n is selected by BAR.

		Sample Size				
Distribution	Kurtosis	50	100	150	200	500
$N(0,1)$	3	0.0602	0.0545	0.0512	0.0516	0.0504
$CN_{0.25,0.4}$	6.4	0.0612	0.0580	0.0570	0.0534	0.0512
t_5	6	0.0618	0.0564	0.0534	0.0502	0.0526
$Laplace$	6	0.0646	0.0572	0.0550	0.0494	0.0526
$NIG_{0.4,0,1,0.6}$	12.5	0.0596	0.0504	0.0482	0.0472	0.0486
$Unif$	1.80	0.0650	0.0536	0.0530	0.0538	0.0504
$Logistic$	1.20	0.0622	0.0580	0.0482	0.0512	0.0496

Table 5.4: Size of the NT-GARCH test for the processes following Model 2 with various heavy-tailed distributions at $\alpha = 5\%$ using $MC = 5000$ and $B = 999$. The window length k_n is selected by BAR.

The improvement of BNT-GARCH over PNT-GARCH on size is substantial in skewed distributions (see Table 5.5 and 5.6). Notice that the sizes of PNT-GARCH fluctuate at a certain level for ARCH processes, but monotonically decrease with a growing n for GARCH processes, e.g., for an ARCH(1) process from an exponentially distribution, the sizes of PNT-GARCH are 0.0284 and 0.0286 respectively for 50 and 500 observations; for a GARCH(1,1) process from a lognormal distribution, PNT-GARCH provide sizes of 0.0564 and 0.0412 respectively for 50 and 500 observations. By contrast, the sizes of BNT-GARCH are quite close to 0.05 even for small and moderate n, e.g., for an ARCH(1) process with exponentially distributed $\{e_t\}$, the sizes of BNT-GARCH are 0.0318 and 0.0358 respectively for 50 and 150 observations; for a GARCH(1,1) process with lognormally distributed $\{e_t\}$, BNT-GARCH had sizes of 0.0466 and 0.0454 respectively for 100 and 200 observations. Therefore, our empirical evidence indicates

that: a rule of thumb of applying the NT-GARCH test is to use the bootstrap procedure to select k_n and then obtain the respective bootstrapped critical value.

				Sample Size				
Distribution	Skewness	Kurtosis	Method	50	100	150	200	500
$Exp(1)$	2	9	PNT-GARCH	0.0284	0.0234	0.0278	0.0260	0.0286
			BNT-GARCH	0.0318	0.0340	0.0358	0.0362	0.0412
$\ln N(0, 0.5)$	2.939	25.507	PNT-GARCH	0.0408	0.0384	0.0384	0.0442	0.0356
			BNT-GARCH	0.0520	0.0468	0.0442	0.0506	0.0432
$NIG_{0.5,0.2,1,0.5}$	2.5	21.5	PNT-GARCH	0.0388	0.0348	0.0334	0.0312	0.0284
			BNT-GARCH	0.0536	0.0414	0.0418	0.0412	0.0426

Table 5.5: Size of the NT-GARCH test for the processes following Model 1 with skewed distributions at $\alpha = 5\%$ using $MC = 5000$ and $B = 999$. The window length k_n is selected by BAR.

				Sample Size				
Distribution	Skewness	Kurtosis	Method	50	100	150	200	500
$Exp(1)$	2	9	PNT-GARCH	0.0430	0.0338	0.0398	0.0322	0.0303
			BNT-GARCH	0.0418	0.0448	0.0444	0.0408	0.0420
$\ln N(0, 0.5)$	2.939	25.507	PNT-GARCH	0.0564	0.0476	0.0408	0.0402	0.0412
			BNT-GARCH	0.0542	0.0466	0.0448	0.0454	0.0444
$NIG_{0.5,0.2,1,0.5}$	2.5	21.5	PNT-GARCH	0.0560	0.0452	0.0404	0.0396	0.0364
			BNT-GARCH	0.0540	0.0472	0.0436	0.0444	0.0456

Table 5.6: Size of the NT-GARCH test for the processes following Model 2 with skewed distributions at $\alpha = 5\%$ using $MC = 5000$ and $B = 999$. The window length k_n is selected by BAR.

5.3.2 Power of the NT-GARCH test

In this section, let us investigate the power of the NT-GARCH test. In particular, we consider the following regression functions:

$$g_1(x) = 64x^3(1-x)^3, \tag{5.38}$$

$$g_2(x) = 0.5\big(1+\sin(3\pi x)\big), \tag{5.39}$$

$$g_3(x) = 0.2x + 0.5\big(1+\sin(3\pi x)\big). \tag{5.40}$$

where $x = t/n$, $t = 1, \ldots, n$, and $\{\epsilon_t\}$ follows an ARCH process defined in Model 1. Note that g_1 and g_2 are discussed in Example 4.1 of Wang et al.(2008), and g_3 is designed to be a monotonic trend mixed with seasonality (see Figure 5.5 for examples of g_1, g_2 and g_3 with $\{\epsilon_t\}$ following an ARCH(1) process).

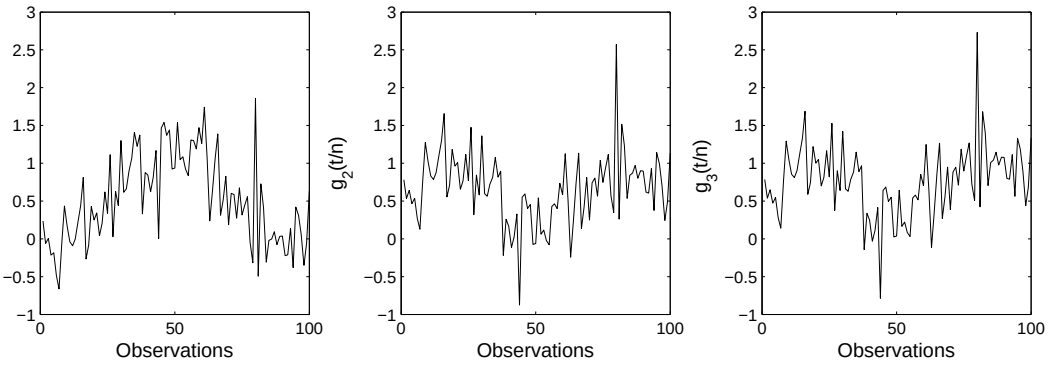

Figure 5.5: Processes of length 100 respectively follow g_1, g_2 and g_3 with ARCH(1) innovations.

Since BNT-GARCH consistently outperforms PNT-GARCH in terms of size, here we only analyze the power of BNT-GARCH with k_n selected by BAR (see **Remark 6** for the results of BNT-GARCH with fixed k_n). In order to demonstrate our result, we compare the power of

BNT-GARCH to that of Mann-Kendall. Table 5.7 shows the power of the BNT-GARCH and the Mann-Kendall tests for processes following respectively g_1, g_2 and g_3 with ARCH(1) innovations. Clearly, since the functional forms of g_1, g_2 and g_3 are non-monotonic, the powers of Mann-Kendall under such alternative hypotheses are close to zero. However, for both heavy-tailed and skewed distributions, BNT-GARCH provides powers extremely close to 1, even in small samples of 50 or 100. Note that we obtain similar results on power when the innovations $\{\epsilon_t\}$ follows a GARCH process. In short, BNT-GARCH is suitable for detecting non-monotonic trends under ARCH/GARCH effect mixed with certain seasonality, where Mann-Kendall typically fails.

			Sample Size	
Distribution	Model	Method	50	100
$N(0,1)$	g_1	Mann-Kendall	0.0206	0.0198
		BNT-GARCH	0.9084	0.9900
	g_2	Mann-Kendall	0.0162	0.0200
		BNT-GARCH	0.8782	0.9852
	g_3	Mann-Kendall	0.0518	0.1102
		BNT-GARCH	0.8932	0.9886
t_5	g_1	Mann-Kendall	0.0206	0.0182
		BNT-GARCH	0.8804	0.9670
	g_2	Mann-Kendall	0.0144	0.0156
		BNT-GARCH	0.8664	0.9492
	g_3	Mann-Kendall	0.0530	0.1164
		BNT-GARCH	0.8728	0.9528
$Unif$	g_1	Mann-Kendall	0.0230	0.0212
		BNT-GARCH	0.9358	0.9986
	g_2	Mann-Kendall	0.0238	0.0244
		BNT-GARCH	0.9122	0.9952
	g_3	Mann-Kendall	0.0558	0.1006
		BNT-GARCH	0.9238	0.9964
$Exp(1)$	g_1	Mann-Kendall	0.0328	0.0312
		BNT-GARCH	0.8518	0.9396
	g_2	Mann-Kendall	0.0014	0.0001
		BNT-GARCH	0.8712	0.9516
	g_3	Mann-Kendall	0.0268	0.0690
		BNT-GARCH	0.8766	0.9556

Table 5.7: Power of the BNT-GARCH and the Mann-Kendall tests at $\alpha = 5\%$ for the processes following g_1, g_2 and g_3 with ARCH(1) innovations using $MC = 5000$ and $B = 999$. The window length k_n is selected by BAR.

Remark 6. The previous simulation studies demonstrate that in small and moderate samples, k_n selected by BAR provides a size closer to the nominal significance level, compared to fixed k. Here we show that the k_n which increases with n actually enables NT-GARCH to gain better power under alternative hypotheses. For example, let us consider the following regression functions:

$$g_2'(x) = 0.5\big(1+\sin(\pi x)\big), \tag{5.41}$$

$$g_3'(x) = 0.2x + 0.5\big(1+\sin(\pi x)\big), \tag{5.42}$$

where $x = t/n$, $t = 1, \ldots, n$, and $\{\epsilon_t\}$ follows an ARCH(1) process. Table 5.8 presents the power of BNT-GARCH when $k_n = 5$ (denoted as BNT-GARCH$_{k_n=5}$), $k_n = 9$ (BNT-GARCH$_{k_n=5}$) and k_n selected by BAR. As k_n increase from 5 to 9, the power of BNT-GARCH improves correspondingly in all cases, but such improvement ceases as k_n gets too large. Therefore, we need to select an "optimal" k_n by BAR, which provides the best power among the three according to the results in Table 5.8.

Distribution	Model	Method	Sample Size	
			50	100
$N(0,1)$	g_2'	BNT-GARCH$_{k_n=5}$	0.4074	0.5944
		BNT-GARCH$_{k_n=9}$	0.4266	0.6884
		BNT-GARCH	0.5245	0.7123
	g_3'	BNT-GARCH$_{k_n=5}$	0.4274	0.6244
		BNT-GARCH$_{k_n=9}$	0.4532	0.6973
		BNT-GARCH	0.5444	0.7662
t_5	g_2'	BNT-GARCH$_{k_n=5}$	0.3642	0.5632
		BNT-GARCH$_{k_n=9}$	0.3996	0.5878
		BNT-GARCH	0.4654	0.6734
	g_3'	BNT-GARCH$_{k_n=5}$	0.3833	0.5932
		BNT-GARCH$_{k_n=9}$	0.4212	0.6854
		BNT-GARCH	0.5006	0.7454

Table 5.8: Power comparisons of the BNT-GARCH when k_n is fixed and selected by BAR at $\alpha = 5\%$ for the processes following g_2' and g_3' with ARCH(1) innovations using $MC = 5000$ and $B = 999$.

5.4 Case Studies

In this section, we apply the proposed NT-GARCH test to detect trend in water and air pollution data. In particular, we study the concentration levels of sulfate anion in the Turkey Lakes Watershed, and of carbon monoxide in the air of Hamilton (ON), Windsor (ON), Montreal (QC), Leth-

bridge (AB). In both examples, we first verify the presence of GARCH effect before applying the trend test and then present the test results by BNT, BNT-GARCH as well as Mann-Kendall. Note that k_n is selected by BAR for BNT-GARCH.

5.4.1 Turkey Lakes Watershed data

Acid rain continues to be a major environmental problem in Canada. Since 1980 Environment Canada, Natural Resources Canada and Fisheries and Oceans Canada have established an intensive research project on Turkey Lakes Watershed (TLW) to define the impact of acidic deposition on undeveloped aquatic and terrestrial terrain. Our analysis focuses on the weekly concentration level of sulfate anion (SO_4^{2-}) in TLW from 1993/01/05 to 2002/12/31, total of 504 observations (see Figure 5.6). We are interested in determining whether a trend is present in SO_4^{2-} level, as it is important for characterizing and controlling anthropogenic acid rain. [2]

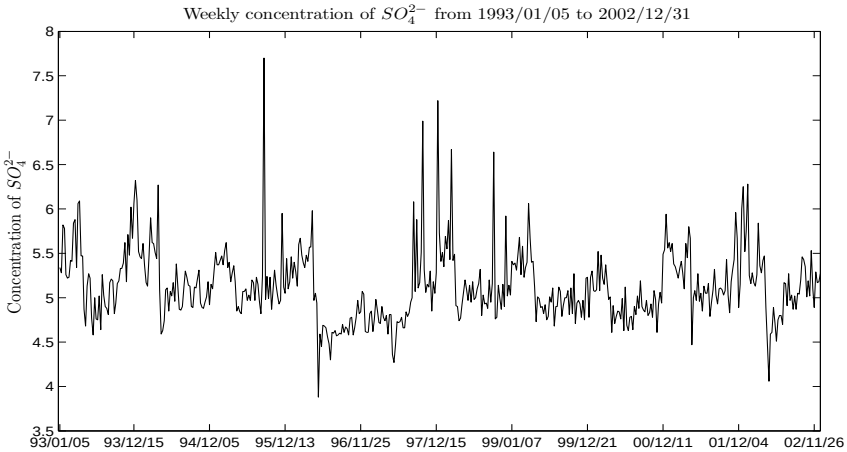

Figure 5.6: Plot of weekly SO_4^{2-} from 1993/01/5 to 2002/12/31.

[2]The Turkey lakes watershed data are kindly provided by Dr. Abdel El-Shaarawi.

Let us denote the SO_4^{2-} data as $\{Y_{t,SO_4^{2-}}\}$. Figure 5.6 indicates that there might be some non-monotonic trend in $\{Y_{t,SO_4^{2-}}\}$. In particular, the mean level of the data considerably decreases between the year of 1995 and 1997, and it exhibits a slight downward trend since 1997. Therefore, we systematically verify our initial observation using the Mann-Kendall, the BNT and the BNT-GARCH tests, respectively. First of all, let us check if there is an ARCH effect in $\{Y_{t,SO_4^{2-}}\}$. Clearly, the data are serially correlated: the Akaike Information Criterion (AIC) suggests an AR(4) model to account for the underlying correlation. Also, notice that $\{Y_{t,SO_4^{2-}}\}$ displays the pattern of seasonal cycle every year, where the mean level largely drops in May, slowly increases during summer and fall, and finally reaches the maximum in winter. Such behavior exactly matches the snow-melting process. Thus, we pre-whiten $\{Y_{t,SO_4^{2-}}\}$ using the following model:

$$Y_{t,SO_4^{2-}} = \varphi_0 + \varphi_1 \sin(2\pi \text{time}) + \varphi_2 \cos(2\pi \text{time}) + v_t, \text{ with } v_t = \sum_{i=1}^{4} \phi_i v_{t-i} + \eta_t, \qquad (5.43)$$

where

$$\text{time} = (\text{year} - 1993) + (\text{month} - 1)/12 + \text{day}/365; \qquad (5.44)$$

η_t is white noise (WN); φ_j, $j = 0, 1, 2$, are some constants; $\phi_i \in (0, 1)$, $i = 1, \ldots, 4$. The LS estimates of parameters of model (5.43) are

$$(\hat{\varphi}_0, \hat{\varphi}_1, \hat{\varphi}_2, \hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3, \hat{\phi}_4) = (5.131, 0.0636, 0.2352, 0.3375, 0.1213, 0.0758, 0.0882),$$

which are all statistically significant at $\alpha = 0.05$. It should be noted that the standard errors of the parameter estimates may be affected by potential departures from the assumptions of normality and independence. However, the corresponding residuals give us an opportunity to test for the ARCH effect. Next, we examine if the ARCH effect is present in the residuals of fitted model $\{\hat{\eta}_t\}$. As shown in Figure 5.7, there is no significant autocorrelations among $\{\hat{\eta}_t\}$ (left panel),

but the correlation at lag 1 is large in the squared-residuals $\{\hat{\eta}_t^2\}$ (the right panel), which is an indication of possible ARCH effect.

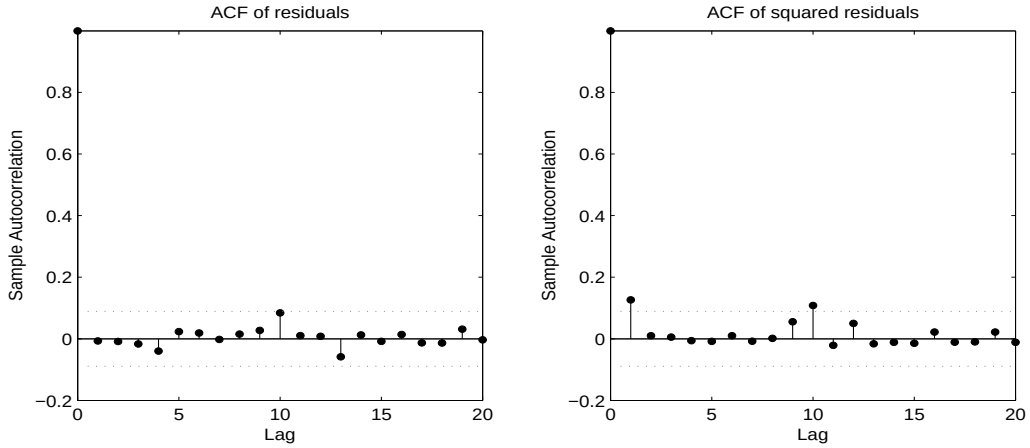

Figure 5.7: The ACF plots of residuals $\{\hat{\eta}_t\}$ and squared residuals $\{\hat{\eta}_t^2\}$ of the fitted model (5.43).

In addition, the p-values of the Lagrange Multiplier (LM), the bootstrap LM (BLM) and the permutation LM (PLM) tests respectively are 0.0098, 0.021 and 0.013, assuming that the ARCH order is 1, i.e., all three tests consistently provide some evidence for ARCH effect.

Next, we test for trend in each of the original $\{Y_{t,SO_4^{2-}}\}$ and the pre-whitened data $\{\hat{\eta}_t\}$. Table 5.9 shows the p-values of Mann-Kendall, BNT and BNT-GARCH applied to the process $\{Y_{t,SO_4^{2-}}\}$. As a result, Mann-Kendall and BNT do not reject H_0 that there is no trend, but BNT-GARCH provides strong evidence (p-value equal to 0.001) against H_0, implying that the SO_4^{2-} data is likely to contain a trend of some functional form.

Test	p-value	k_n	Model
Mann-Kendall	0.1088		
BNT	0.5876	5	AR(4)
BNT-GARCH	0.0010	39	ARCH(1)

Table 5.9: The p-values of Mann-Kendall, BNT and BNT-GARCH applied to $\{Y_{t,SO_4^{2-}}\}$. The BNT test is based on an AR(4) model and the window length $k_n = 5$, while the BNT-GARCH test assume an ARCH(1) model and $k_n = 39$. The number of bootstrap replication is 999.

Since it is well-known that Mann-Kendall has inflated Type I error under serial correlation (Hamed and Rao, 1998; Hamed, 2009; Noguchi et al., 2011) or seasonality (EI-Shaarawi and Niculescu, 1992), we also provide an analysis based on the pre-whitened data $\{\hat{\eta}_t\}$ in order to confirm our test results. Note that the ARCH effect still remains in $\{\hat{\eta}_t\}$, with p-values of LM, BLM and PLM to be 0.0356, 0.041 and 0.028 respectively. Next, we apply the Mann-Kendall, BNT and BNT-GARCH tests, respectively, to $\{\hat{\eta}_t\}$. As the p-values show in Table 5.10, Mann-Kendall and BNT do not reject H_0, while BNT-GARCH provides some evidence (p-value equal to 0.0441) against H_0, indicating that trend may exist in the data at significance level $\alpha = 0.05$. Hence, the new results based on $\{\hat{\eta}_t\}$ are consistent with the previous ones.

Test	p-value	k_n	Model
Mann-Kendall	0.6946		
BNT	0.6256	5	WN
BNT-GARCH	0.0441	39	ARCH(1)

Table 5.10: The p-values of Mann-Kendall, BNT and BNT-GARCH applied to $\{\hat{\eta}_t\}$, after removing the serial correlation and seasonality from $\{Y_{t,SO_4^{2-}}\}$. The BNT test uses window length $k_n = 5$, assuming $\{\hat{\eta}_t\}$ is white noise, and the BNT-GARCH test is based on an ARCH(1) model and $k_n = 39$. The number of bootstrap replication is 999.

In summary, given that the original data $\{Y_{t,SO_4^{2-}}\}$ display some cyclic patterns in the mean level and there is evidence of ARCH effect, we proceed with the findings of the BNT-GARCH test, i.e., we reject H_0 that no trend is in $\{Y_{t,SO_4^{2-}}\}$ at significance level $\alpha = 0.05$.

5.4.2 Air pollution data

Although air pollution in Canada is not as severe as many other countries in the world, it is still among the major environmental problems that Canadians are facing, e.g., the smog (a combination of smoke and fog) event along the Windsor-Quebec City corridor of Ontario and Quebec. The two main contaminants of smog are ground level ozone and small airborne particles, such as nitrogen oxides, sulfur dioxide and carbon monoxide (CO). Typically, the smog levels peak between May and September. We would like to verify if this phenomenon can be reflected in the concentration level of CO over time, i.e., our goal is to test if there is any trend in the CO level. [3]

[3]The data are downloaded from the NAPS database of Environment Canada.

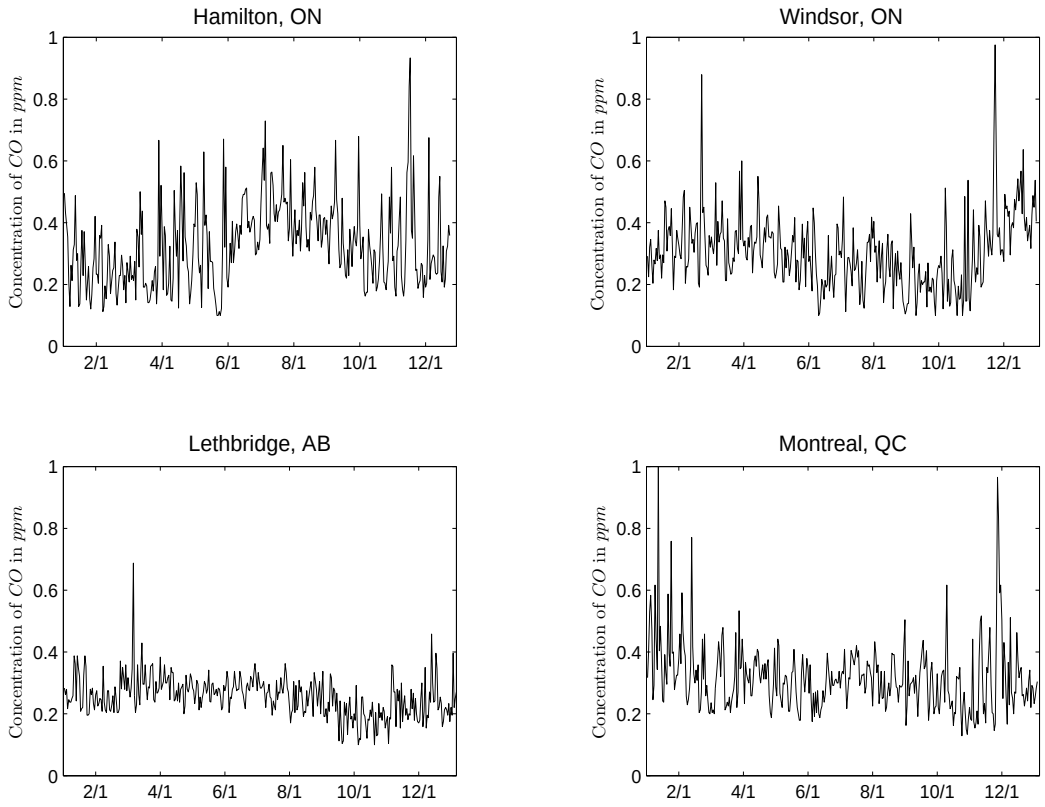

Figure 5.8: The concentration of CO in 2006 respectively in Hamilton, Windsor, Montreal and Lethbridge.

In our study, we select four different sites including Hamilton (ON), Windsor (ON), Montreal (QC), Lethbridge (AB), three of which are along the Windsor-Quebec City corridor but one of which is not. Our data sets include the daily averages of the concentrations of CO in ppm from 2006/1/1 to 2006/12/31 on these four sites, respectively denoted as $\{Y_{t,H}\}$, $\{Y_{t,W}\}$, $\{Y_{t,M}\}$ and $\{Y_{t,L}\}$. Figure 5.8 presents the observed data. $\{Y_{t,H}\}$ shows a clear upward trend since the end

of May, steadily reaches a peak in July and then declines afterwards. Surprisingly, there is a decrease in the mean level of $\{Y_{t,W}\}$ over the summer while no clear trend is shown in $\{Y_{t,M}\}$. The mean level of $\{Y_{t,L}\}$ is lower than those of $\{Y_{t,H}\}$, $\{Y_{t,W}\}$ and $\{Y_{t,M}\}$, with a decrease between mid-September to mid-November.

Similar to the previous case study, we first verify if the ARCH assumption is valid for the CO processes. As the data are temporally dependent, we pre-whiten each data set by an AR model whose order is selected by AIC. Let us denote the pre-whitened data as $\{\tilde{Y}_{t,H}\}$, $\{\tilde{Y}_{t,W}\}$, $\{\tilde{Y}_{t,M}\}$ and $\{\tilde{Y}_{t,L}\}$. Then the BLM and PLM tests are applied to $\{\tilde{Y}_{t,H}\}$, $\{\tilde{Y}_{t,W}\}$, $\{\tilde{Y}_{t,M}\}$ and $\{\tilde{Y}_{t,L}\}$ in order to detect the ARCH effect (see the test results in Table 5.11 and the ACF plots of the squared residuals in Figure 5.9). All the p-values of BLM and PLM are highly statistically significant. Hence, there is strong evidence for ARCH effects in $\{Y_{t,H}\}$, $\{Y_{t,W}\}$, $\{Y_{t,M}\}$ and $\{Y_{t,L}\}$.

		p-values	
Data	Model	BLM	PLM
$\{Y_{t,H}\}$	AR(1)	0	0.001
$\{Y_{t,W}\}$	AR(5)	0.003	0.005
$\{Y_{t,M}\}$	AR(7)	0.033	0.023
$\{Y_{t,L}\}$	AR(8)	0.013	0.016

Table 5.11: The BLM and PLM p-values of the residuals of $\{Y_{t,H}\}$, $\{Y_{t,W}\}$, $\{Y_{t,M}\}$ and $\{Y_{t,L}\}$ after fitting AR(1), AR(5), AR(7) and AR(8), respectively.

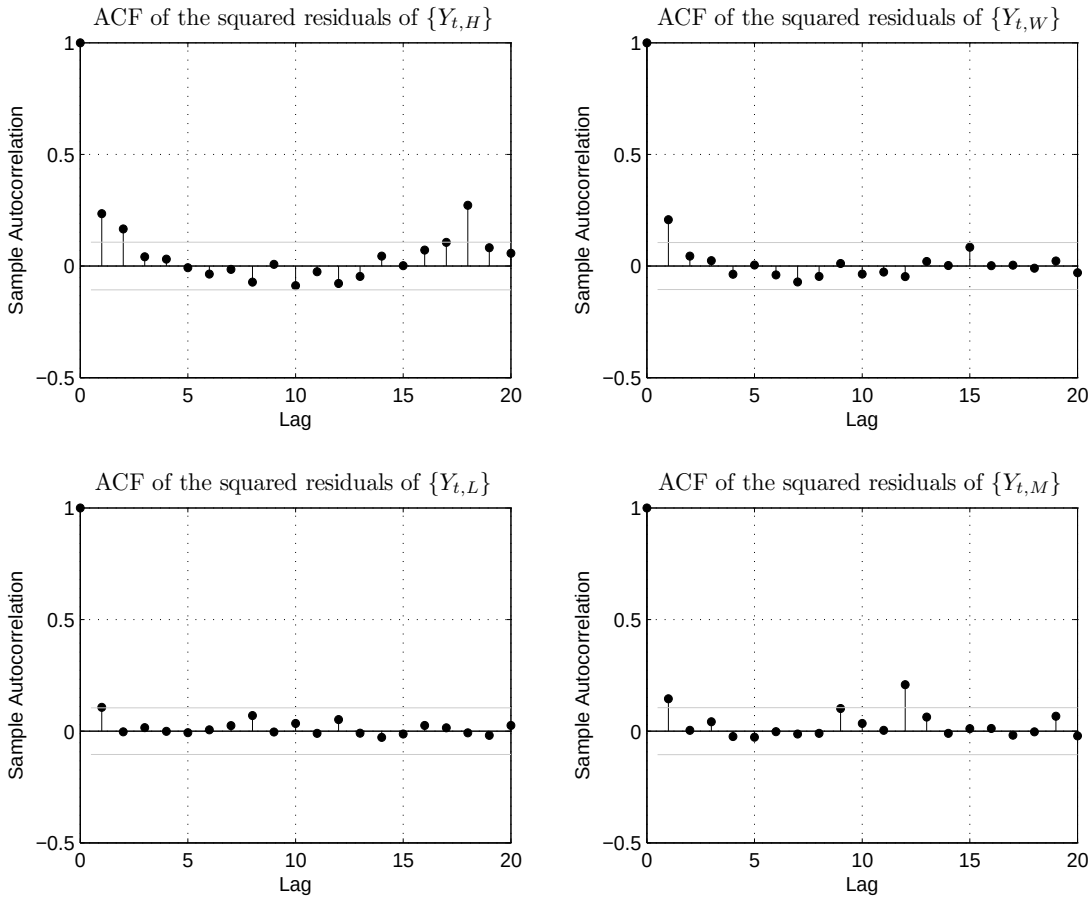

Figure 5.9: The ACF plots of the squared residuals of $\{Y_{t,H}\}$, $\{Y_{t,W}\}$, $\{Y_{t,M}\}$ and $\{Y_{t,L}\}$ after fitting AR(1), AR(5), AR(7) and AR(8), respectively.

Next, we apply the Mann-Kendall and BNT-GARCH tests to detect possible trends in $\{Y_{t,H}\}$, $\{Y_{t,W}\}$, $\{Y_{t,M}\}$ and $\{Y_{t,L}\}$ (BNT does not detect any non-monotonic trend under ARCH effect according to our simulation results and the previous case study, and hence we omit here). Note that we assume ARCH(1) effect for BNT-GARCH in all cases. Table 5.12 shows p-values of

Mann-Kendall and BNT-GARCH of the original processes $\{Y_{t,H}\}$, $\{Y_{t,W}\}$, $\{Y_{t,M}\}$ and $\{Y_{t,L}\}$. Since all p-values are highly significant, there is strong evidence indicating the presence of trends in $\{Y_{t,H}\}$, $\{Y_{t,W}\}$, $\{Y_{t,M}\}$ and $\{Y_{t,L}\}$ by both Mann-Kendall and BNT-GARCH.

Data	Test	p-values	k_n
$\{Y_{t,H}\}$	Mann-Kendall	0.0013	
	BNT-GARCH	0	35
$\{Y_{t,W}\}$	Mann-Kendall	0.0198	
	BNT-GARCH	0.001	33
$\{Y_{t,M}\}$	Mann-Kendall	0	
	BNT-GARCH	0	33
$\{Y_{t,L}\}$	Mann-Kendall	0.00004	
	BNT-GARCH	0.002	35

Table 5.12: The p-values of the Mann-Kendall and the BNT-GARCH tests applied to $\{Y_{t,H}\}$, $\{Y_{t,W}\}$ $\{Y_{t,M}\}$ $\{Y_{t,L}\}$. The number of bootstrap replications is 999.

However, according to Figure 5.8, the concentration level of CO in Montreal ($\{Y_{t,M}\}$) does not show a clear change in mean, which makes the previous result quite doubtful. Hence, we again apply Mann-Kendall and BNT-GARCH to the pre-whitened processes $\{\hat{\eta}_{t,H}\}$, $\{\hat{\eta}_{t,W}\}$, $\{\hat{\eta}_{t,M}\}$ and $\{\hat{\eta}_{t,L}\}$, by fitting $AR(1)$, $AR(5)$, $AR(7)$ and $AR(8)$ respectively to $\{Y_{t,H}\}$, $\{Y_{t,W}\}$, $\{Y_{t,M}\}$ and $\{Y_{t,L}\}$ (see Table 5.13). After removing the correlation in the data, the p-values of Mann-Kendall and BNT-GARCH for Montreal ($\{\hat{\eta}_{t,M}\}$) increase respectively to 0.0785 and 0.2352, showing no evidence against the null hypothesis at significance level $\alpha = 0.05$. Moreover, BNT-GARCH indicates no trend in $\{\hat{\eta}_{t,M}\}$ even at $\alpha = 0.1$. A possible explanation could

be that $\{Y_{t,M}\}$ is collected at a station near Montreal airport, where the CO level is quite stable during a year. Also, there is a discrepancy between Mann-Kendall and BNT-GARCH for Windsor ($\{\hat{\eta}_{t,W}\}$), where the p-value of Mann-Kendall is 0.6953, not rejecting H_0 of no trend, but that of BNT-GARCH is 0.036, showing some evidence of trend. Since there is a clear shift in the mean level of $\{Y_{t,W}\}$ by Figure 5.8, we incline to accept the conclusion by BNT-GARCH.

Data	Test	p-values	k_n
$\{\hat{\eta}_{t,H}\}$	Mann-Kendall	0.0479	
	BNT-GARCH	0.001	47
$\{\hat{\eta}_{t,W}\}$	Mann-Kendall	0.6953	
	BNT-GARCH	0.0360	41
$\{\hat{\eta}_{t,M}\}$	Mann-Kendall	0.0785	
	BNT-GARCH	0.2352	45
$\{\hat{\eta}_{t,L}\}$	Mann-Kendall	0.0343	
	BNT-GARCH	0.048	47

Table 5.13: The p-values of the Mann-Kendall, BNT and NT tests applied to pre-whitened processes $\{\hat{\eta}_{t,H}\}$, $\{\hat{\eta}_{t,W}\}$, $\{\hat{\eta}_{t,M}\}$ and $\{\hat{\eta}_{t,L}\}$. The number of bootstrap replications is 999.

Appendix of Chapter 5

We first state a lemma about moments of GARCH processes, which is one of the fundamental arguments to show that T_n is asymptotically normally distributed.

Lemma A.1 *Suppose that H_0 holds. Let $p_1, \ldots, p_l \in \mathbb{N} \cup \{0\}$ and $i_1 < i_2 < \ldots < i_l$. If e_t*

is symmetrically distributed and there exists at least one $s \in \{1, \ldots, l\}$ *such that* p_s *is an odd number, then*

$$E\big(Y_{i_1}^{p_1} \ldots Y_{i_l}^{p_l}\big) = 0.$$

Proof of Lemma A.1 Suppose that p_s is an odd number. Note that $\sigma_{i_1}^{p_1} e_{i_1}^{p_1} \ldots \sigma_{i_s}^{p_s}$ is independent from e_{i_s}, and $\sigma_{i_{s+1}}^{p_{s+1}} e_{i_{s+1}}^{p_{s+1}} \ldots \sigma_{i_l}^{p_l} e_{i_l}^{p_l}$ depends on e_{i_s} only through $e_{i_s}^2$, i.e., $\sigma_{i_{s+1}}^{p_{s+1}} e_{i_{s+1}}^{p_{s+1}} \ldots \sigma_{i_l}^{p_l} e_{i_l}^{p_l}$ conditional on e_{i_s} is equal to $\sigma_{i_{s+1}}^{p_{s+1}} e_{i_{s+1}}^{p_{s+1}} \ldots \sigma_{i_l}^{p_l} e_{i_l}^{p_l}$ conditional on $-e_{i_s}$. Therefore, using the symmetry of e_t, we obtain

$$\begin{aligned} E\big(Y_{i_1}^{p_1} \ldots Y_{i_s}^{p_s} \ldots Y_{i_l}^{p_l}\big) &= E\big(\sigma_{i_1}^{p_1} e_{i_1}^{p_1} \ldots \sigma_{i_s}^{p_s} e_{i_s}^{p_s} \ldots \sigma_{i_l}^{p_l} e_{i_l}^{p_l}\big) \\ &= E\big(\sigma_{i_1}^{p_1} e_{i_1}^{p_1} \ldots \sigma_{i_s}^{p_s} (-e_{i_s})^{p_s} \ldots \sigma_{i_l}^{p_l} e_{i_l}^{p_l}\big) \\ &= -E\big(\sigma_{i_1}^{p_1} e_{i_1}^{p_1} \ldots \sigma_{i_s}^{p_s} e_{i_s}^{p_s} \ldots \sigma_{i_l}^{p_l} e_{i_l}^{p_l}\big) \\ &= -E\big(Y_{i_1}^{p_1} \ldots Y_{i_s}^{p_s} \ldots Y_{i_l}^{p_l}\big) \end{aligned} \tag{5.45}$$

and the result follows. □

Proof of Lemma 5.1.1 By definition, $\boldsymbol{A}$ and $\boldsymbol{A}_d$ are $nk_n \times nk_n$ contrast matrices containing $n \times n$ blocks, each of size $k_n \times k_n$. Let (i_1, i_2) denote the row and column indices of blocks within $\boldsymbol{A}$ and $\boldsymbol{A}_d$, $i_1,\ i_2 = 1, \ldots, n$. Notice that $\boldsymbol{A}$ and $\boldsymbol{A}_d$ have exactly the same elements in the diagonal blocks ($i_1 = i_2$) and hence the diagonal blocks of $(\boldsymbol{A} - \boldsymbol{A}_d)$ are identically zero. For $i_1 \neq i_2$, every element in the block (i_1, i_2) of $(\boldsymbol{A} - \boldsymbol{A}_d)$ is equal to $-1/n(n-1)k_n$. To prove this lemma it is sufficient to show that both the expectation and the variance of $\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A} - \boldsymbol{A}_d)\boldsymbol{V}$ tend to 0 as $n \to \infty$. Note that for a given sample $\{Y_1, \ldots, Y_n\}$, there are $n - k_n + 1$ symmetric windows, that is, $\{W_{(k_n+1)/2}, \ldots, W_{n-(k_n-1)/2}\}$. Let us first analyze the expectation. Because

$\{Y_t, t \in \mathbb{Z}\}$ is an uncorrelated process, we obtain

$$\begin{aligned} & E\big(\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{A}_d)\boldsymbol{V}\big) \\ = & -\frac{1}{\sqrt{n}(n-1)k_n^{3/2}} \sum_{\substack{i_1,i_2=(k_n+1)/2 \\ i_1\neq i_2}}^{n-(k_n-1)/2} \sum_{j_1,j_2=1}^{n} E(Y_{j_1}Y_{j_2}) I(j_1 \in W_{i_1}, j_2 \in W_{i_2}) \\ = & -\frac{1}{\sqrt{n}(n-1)k_n^{3/2}} \sum_{\substack{i_1,i_2=(k_n+1)/2 \\ i_1\neq i_2}}^{n-(k_n-1)/2} \sum_{j=1}^{n} E(Y_j^2) I(j \in W_{i_1} \cap W_{i_2}). \end{aligned} \tag{5.46}$$

For every $k_n \leq j \leq n-k_n+1$ on the right hand side of (5.46), there are exactly k_n^2 combinations of i_1, i_2 for which $j \in W_{i_1} \cap W_{i_2}$, namely, $i_1, i_2 \in \{j-(k_n-1)/2, \ldots, j+(k_n-1)/2\}$. In the boundary cases when $j < k_n$ or $j > n-k_n+1$, there are j^2 and $(n-j)^2$ combinations, respectively. Thus, using the stationarity of $\{Y_t,\ t \in \mathbb{Z}\}$, equation (5.46) becomes

$$\begin{aligned} E\big(\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{A}_d)\boldsymbol{V}\big) &= -\frac{E(Y_t^2)}{\sqrt{n}(n-1)k_n^{3/2}}\Big((n-2k_n+2)k_n^2 + 2\sum_{j=1}^{k_n-1} j^2\Big) \\ &= O\big(n^{-3/2}k_n^{-3/2}(nk_n^2+k_n^3)\big) = O(n^{-1/2}k_n^{1/2}) \\ &= o(1) \end{aligned} \tag{5.47}$$

as $k_n/n \to 0$. For the second moment of $\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{A}_d)\boldsymbol{V}$, we obtain

$$\begin{aligned} & E\big(\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{A}_d)\boldsymbol{V}\big)^2 \\ & \quad = \frac{1}{n(n-1)^2k_n^3} \sum_{\substack{i_1,i_2,i_3,i_4=(k_n+1)/2 \\ i_1\neq i_2, i_3\neq i_4}}^{n-(k_n-1)/2} \sum_{j_1,\ldots,j_4=1}^{n} E(Y_{j_1}Y_{j_2}Y_{j_3}Y_{j_4}) I(j_s \in W_{i_s},\ s=1,\ldots,4). \end{aligned} \tag{5.48}$$

Because of the assumption $E(e_t^4) < \infty$, the expected values on the right-hand side of equation (5.48) are all finite and bounded by $E(Y_t^4)$. Furthermore, according to Lemma A.1, the

expected value is nonzero only if two pairs of (j_1, j_2, j_3, j_4) are equal, i.e., $j_1 = j_2$ and $j_3 = j_4$, or $j_1 = j_3$ and $j_2 = j_4$, or $j_1 = j_4$ and $j_2 = j_3$, which in particular includes the case $j_1 = j_2 = j_3 = j_4$. Consequently, we can re-write equation (5.48) as

$$\begin{aligned} & E\big(\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{A}_d)\boldsymbol{V}\big)^2 \\ = \; & \frac{1}{n(n-1)^2k_n^3} \sum_{\substack{i_1,i_2,i_3,i_4=(k_n+1)/2 \\ i_1\neq i_2, i_3\neq i_4}}^{n-(k_n-1)/2} \sum_{j_1,\ldots,j_4=1}^{n} E(Y_{j_1}Y_{j_2}Y_{j_3}Y_{j_4})\Big\{I(j_s \in W_{i_s},\ s=1,\ldots,4) \\ \times \; & \Big(I(j_1=j_2,\ j_3=j_4)+I(j_1=j_3,\ j_2=j_4)+I(j_1=j_4,\ j_2=j_3)-2I(j_1=j_2=j_3=j_4)\Big)\Big\}. \end{aligned} \tag{5.49}$$

Let us count the number of nonzero terms on the right hand side of (5.49). Clearly, there are n^2 combinations of j_1, j_2, j_3, j_4 satisfying $j_1 = j_2$ and $j_3 = j_4$. For each combination, there are $O(k_n^4)$ ways to choose i_1, i_2, i_3, i_4 such that $j_s \in W_{i_s}$ for $s = 1, \ldots, 4$. Hence, the total number of combinations of j_1, j_2, j_3, j_4 and i_1, i_2, i_3, i_4 satisfying

$$j_1 = j_2,\ j_3 = j_4 \text{ and } j_s \in W_{i_s} \text{ for } s = 1, \ldots, 4$$

is equal to $O(n^2k_n^4)$. The same arguments hold for the other two cases when $j_1 = j_3$ and $j_2 = j_4$, or $j_1 = j_4$ and $j_2 = j_3$. Also, the total number of combinations of j_1, j_2, j_3, j_4 and i_1, i_2, i_3, i_4 satisfying

$$j_1 = j_2 = j_3 = j_4 \text{ and } j_s \in W_{i_s} \text{ for } s = 1, \ldots, 4$$

is equal to $O(nk_n^4)$, as there are n combinations of j_1, j_2, j_3, j_4 where $j_1 = j_2 = j_3 = j_4$ and $O(k_n^4)$ ways to choose i_1, i_2, i_3, i_4 such that $j_s \in W_{i_s}$ for $s = 1, \ldots, 4$. Therefore, replacing $E(Y_{j_1}Y_{j_2}Y_{j_3}Y_{j_4})$ by the upper bound $E(Y_t^4)$, equation (5.49) becomes

$$\begin{aligned} E\big(\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{A}_d)\boldsymbol{V}\big)^2 &= O\big(n^{-3}k_n^{-3}(n^2k_n^4+nk_n^4)\big) \\ &= O(n^{-3}k_n^{-3}n^2k_n^4) = O(n^{-1}k_n) = o(1), \end{aligned} \tag{5.50}$$

as $k_n/n \to 0$ and the result of this lemma follows. $\square$

Remark 7. Note that, in general, the assumption that e_t is symmetrically distributed cannot be relaxed. In particular, if the distribution of e_t is asymmetric, a statement analogous to Lemma A.1 fails to be valid and, as a consequence, the expectation on the right-hand side of equation (5.48) is nonzero also if $j_1 < j_2 = j_3 = j_4$ or if $j_1, j_2 < j_3 = j_4$. Since

$$\sum_{\substack{i_1,i_2=1\\ i_1\neq i_2}}^{n}\sum_{\substack{i_3,i_4=1\\ i_3\neq i_4}}^{n}\sum_{j_1,\ldots,j_4=1}^{n} I(j_1,j_2<j_3=j_4)\, I(j_s\in W_{i_s},\ s=1,\ldots,4) \;\;=\;\; O(n^3k_n^4),$$

we obtain $E\big(\sqrt{n/k_n}\boldsymbol{V}'(\boldsymbol{A}-\boldsymbol{A}_d)\boldsymbol{V}\big)^2 \sim O(n^{-3}k_n^{-3})O(n^3k_n^4) = O(k_n)$, which does not tend to 0 unless $k_n \to 0$ as $n \to \infty$.

Proof of Theorem 5.1.2 By definition, we have

$$\sqrt{\frac{n}{k_n}}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V} = \frac{1}{\sqrt{nk_n}(k_n-1)}\sum_{i=(k_n+1)/2}^{n-(k_n-1)/2}\sum_{\substack{j_1,j_2=1\\ j_1\neq j_2}}^{n} Y_{j_1}Y_{j_2}I(j_1,j_2\in W_i). \tag{5.51}$$

Since $\{Y_t,\ t\in\mathbb{Z}\}$ is an uncorrelated process, we obtain $E(\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}) = 0$ and hence

$$\begin{aligned}&\mathrm{Var}(\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V})\\ =\;& \frac{1}{nk_n(k_n-1)^2}\sum_{i_1,i_2=(k_n+1)/2}^{n-(k_n-1)/2}\sum_{\substack{j_1,l_1,j_2,l_2=1\\ j_1\neq l_1,j_2\neq l_2}}^{n} E(Y_{j_1}Y_{l_1}Y_{j_2}Y_{l_2})I(j_1,l_1\in W_{i_1},j_2,l_2\in W_{i_2}).\end{aligned} \tag{5.52}$$

By Lemma A.1, the right-hand side of equation (5.52) is nonzero only if two pairs of indices j_1, j_2, l_1, l_2 are equal, i.e., $j_1 = j_2$ and $l_1 = l_2$, or $j_1 = l_2$ and $j_2 = l_1$. Hence, we can re-write equation (5.52) as

$$\mathrm{Var}(\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}) \;\;=\;\; \frac{2}{nk_n(k_n-1)^2}\sum_{i_1,i_2=(k_n+1)/2}^{n-(k_n-1)/2}\sum_{\substack{j,l=1\\ j\neq l}}^{n} E(Y_j^2Y_l^2)I(j,l\in W_{i_1}\cap W_{i_2}). \tag{5.53}$$

Let us count the number of non-zero terms on the right hand side of (5.53). For any $j, l \in \{1, 2, \ldots, n\}$ with $j \neq l$, define the set

$$S_{jl} \;=\; \big\{(k_n+1)/2 \le i \le n-(k_n-1)/2 : \quad j,l \in W_i\big\}. \tag{5.54}$$

Writing $|S_{jl}|$ for the cardinality of S_{jl}, we obtain

$$\mathrm{Var}(\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}) \;=\; \frac{2}{nk_n(k_n-1)^2}\sum_{\substack{j,l=1\\ j\neq l}}^{n} |S_{jl}|^2\,\gamma(|j-l|) \tag{5.55}$$

with $\gamma(|j-l|)$ as defined before the theorem. Now, let us explicitely determine the cardinality of S_{jl}. Note that a necessary and sufficient condition for $j, l \in W_i$ is that $i \ge \max(j,l) - (k_n-1)/2$ and $i \le \min(j,l) + (k_n-1)/2$, which allows us to write

$$\begin{aligned} S_{jl} \;=\; &\big\{(k_n+1)/2, \ldots, n-(k_n-1)/2\big\} \\ &\cap \big\{\max(j,l) - (k_n-1)/2, \ldots, \min(j,l) + (k_n-1)/2\big\}. \end{aligned} \tag{5.56}$$

From (5.56) it is obvious that $|S_{jl}| = 0$ if $|j-l| \ge k_n$, therefore, we may assume $|j-l| < k_n$ in the following. Furthermore, it is not difficult to derive the following properties:

$$|S_{jl}| \;=\; \begin{cases} k_n - |j-l| & \text{if } \max(j,l) \ge k_n \text{ and } \min(j,l) \le n-k_n+1, \\ \min(j,l) & \text{if } \max(j,l) < k_n, \\ n - \max(j,l) + 1 & \text{if } \min(j,l) > n-k_n+1. \end{cases} \tag{5.57}$$

(Note that, under the assumptions $|j-l| < k_n$ and $n \ge 2k_n$, it is impossible that both $\max(j,l) < k_n$ and $\min(j,l) > n-k_n+1$.) Now we rewrite (5.55) in the following way:

$$\mathrm{Var}(\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}) \;=\; \frac{4}{nk_n(k_n-1)^2}\sum_{h=1}^{k_n-1}\gamma(h)\sum_{j=1}^{n-h}|S_{j,j+h}|^2. \tag{5.58}$$

Using the properties derived in (5.57), we obtain

$$\begin{aligned}\sum_{j=1}^{n-h}|S_{j,j+h}|^2 &= \sum_{j=k_n-h}^{n-k_n+1}|S_{j,j+h}|^2+\sum_{j=1}^{k_n-h-1}|S_{j,j+h}|^2+\sum_{j=n-k_n+2}^{n-h}|S_{j,j+h}|^2 \\ &= (n-2(k_n-1)+h)(k_n-h)^2+2\sum_{j=1}^{k_n-h-1}j^2 \end{aligned} \tag{5.59}$$

for $h = 1, 2, \ldots, k_n - 1$. Using the fact that the right hand side of (5.59) is asymptotically equivalent to $n(k_n - h)^2$ and inserting this expression back into (5.58) gives us

$$\text{Var}(\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}) \sim \frac{4}{k_n(k_n-1)^2}\sum_{h=1}^{k_n-1}(k_n-h)^2\gamma(h). \tag{5.60}$$

In the case where $k_n = k$ is fixed, (5.60) immediately gives us the expression in (**??**). Next, let us consider the case when $k_n \to \infty$. Note that the second moment of $\{Y_t, t \in \mathbb{Z}\}$ is

$$E(Y_t^2) = \alpha_0\bigg(1-\sum_{i=1}^{p}\alpha_i-\sum_{j=1}^{q}\beta_j\bigg)^{-1} \tag{5.61}$$

(see, e.g., Theorem 2.5 of Francq and Zakoian, 2010). Since $\{Y_t^2, t \in \mathbb{Z}\}$ has a causal ARMA(p, q) representation (Tsay, 2002; Lindner, 2009), by the property of an ARMA(p, q) process, we obtain

$$\gamma(k_n) \to \big(E(Y_t^2)\big)^2 = \alpha_0^2\bigg(1-\sum_{i=1}^{p}\alpha_i-\sum_{j=1}^{q}\beta_j\bigg)^{-2}, \tag{5.62}$$

at exponential rate as $k_n \to \infty$ (see details in Section 3.3 of Brockwell and Davis, 1991). Also, we have

$$\frac{4}{k_n(k_n-1)^2}\underbrace{\sum_{h=1}^{k_n-1}(k_n-h)^2}_{=(k_n-1)k_n(2k_n-1)/6} \sim \frac{4}{k_n^3}\frac{2k_n^3}{6} \sim 4/3. \tag{5.63}$$

By results (5.62) and (5.63), as n and $k_n \to \infty$,

$$\text{Var}(\sqrt{n/k_n}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V}) \to \frac{4}{3}\alpha_0^2\bigg(1-\sum_{i=1}^{p}\alpha_i-\sum_{j=1}^{q}\beta_j\bigg)^{-2}. \tag{5.64}$$

Hence, the result of this theorem follows. □

Next, we are going to establish the asymptotic normality of the statistic T_n when k_n is fixed. A key to the proof is the following representation:

$$\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V} \;=\; n^{-1}\sum_{i=(k+1)/2}^{n-(k-1)/2} A_i \tag{5.65}$$

where

$$A_i \;=\; \frac{1}{k-1}\sum_{\substack{j_1,j_2=i-(k-1)/2\\ j_1\neq j_2}}^{i+(k-1)/2} Y_{j_1}Y_{j_2}. \tag{5.66}$$

Lemma A.2 *Under H_0, the sequence $\{A_i,\, i\in\mathbb{Z}\}$ is strictly stationary and satisfies the strong mixing condition with the mixing coefficients $\tilde{\alpha}(\tau)$ decaying with geometric rate. Moreover, $E(A_i) = 0$ and, under the additional assumptions that e_t is symmetrically distributed and $E(e_t^4)<\infty$,*

$$E(A_iA_{i+m}) \;=\; \frac{4}{(k-1)^2}\sum_{h=1}^{k-1-m}(k-m-h)\gamma(h). \tag{5.67}$$

for $m\in\mathbb{N}_0$.

Proof of Lemma A.2 Since under H_0 the process $\{Y_t,\, t\in\mathbb{Z}\}$ is strictly stationary, the strict stationarity of $\{A_i,\, i\in\mathbb{Z}\}$ is obvious. Let us establish the statement on the mixing coefficient. According to Theorem 8 of Lindner (2009), under H_0 the process $\{Y_t,\, t\in\mathbb{Z}\}$ is strongly mixing with geometric rate. Let us recall the definition of strongly mixing: for $t\in\mathbb{Z}$ write $\mathfrak{M}_{-\infty}^{t}$ and $\mathfrak{M}_{t}^{\infty}$ to denote the sigma-fields generated by the sequences $\{\ldots,Y_{t-1},Y_t\}$ and $\{Y_t,Y_{t+1},\ldots\}$, respectively. For $\tau\in\mathbb{N}$ define

$$\alpha(\tau) \;=\; \sup_{A\in\mathfrak{M}_{-\infty}^{0},B\in\mathfrak{M}_{\tau}^{\infty}} \big|P(A\cap B)-P(A)P(B)\big|. \tag{5.68}$$

Strong mixing of $\{Y_t,\, t \in \mathbb{Z}\}$ with geometric rate means that there exists some $0 < \kappa < 1$ with

$$\alpha(\tau) \;\; = \;\; O(\kappa^\tau). \tag{5.69}$$

Now let $\tilde{\mathfrak{M}}_{-\infty}^t$ and $\tilde{\mathfrak{M}}_t^\infty$ denote the sigma-fields generated by $\{\ldots, A_{t-1}, A_t\}$ and $\{A_t, A_{t+1}, \ldots\}$, respectively, and define

$$\tilde{\alpha}(\tau) \;\; = \;\; \sup_{A \in \tilde{\mathfrak{M}}_{-\infty}^0,\, B \in \tilde{\mathfrak{M}}_\tau^\infty} \big|P(A \cap B) - P(A)P(B)\big|. \tag{5.70}$$

Note that for every $t \in \mathbb{Z}$, the sequence $\{\ldots, A_{t-1}, A_t\}$ is measurable with respect to the sigma-algebra generated by $\{\ldots, Y_{t+(k-1)/2-1}, Y_{t+(k-1)/2}\}$, and $\{A_t, A_{t+1}, \ldots\}$ is measurable with respect to the sigma-algebra generated by $\{Y_{t-(k-1)/2}, Y_{t-(k-1)/2+1}, \ldots\}$. Consequently,

$$\tilde{\mathfrak{M}}_{-\infty}^0 \;\subset\; \mathfrak{M}_{-\infty}^{(k-1)/2} \quad \text{and} \quad \tilde{\mathfrak{M}}_\tau^\infty \;\subset\; \mathfrak{M}_{\tau-(k-1)/2}^\infty, \tag{5.71}$$

and hence, for $\tau \geq k$,

$$\tilde{\alpha}(\tau) \;\; \leq \sup_{A \in \mathfrak{M}_{-\infty}^{(k-1)/2},\, B \in \mathfrak{M}_{\tau-(k-1)/2}^\infty} \big|P(A \cap B) - P(A)P(B)\big| \;=\; \alpha(\tau - k + 1). \tag{5.72}$$

Putting together (5.69) and (5.72), we obtain $\tilde{\alpha}(\tau) = O(\kappa^{\tau-k+1}) = O(\kappa^\tau)$, which proves the result on the mixing coefficients.

The statement $E(A_i) = 0$ follows from the fact that under H_0 the process $\{Y_t,\, t \in \mathbb{Z}\}$ is uncorrelated. To establish the statement on the second moments, note that

$$E(A_i A_{i+m}) \;\; = \;\; \frac{4}{(k-1)^2} \sum_{\substack{j_1, j_2, l_1, l_2 = 1 \\ j_1 < j_2, l_1 < l_2}}^{k} E(Y_{j_1} Y_{j_2} Y_{l_1+m} Y_{l_2+m}). \tag{5.73}$$

According to Lemma A.1, a necessary condition for the expectation on the right hand side of (5.73) to be nonzero is that $j_1 = l_1 + m$ and $j_2 = l_2 + m$, which gives us

$$E(A_i A_{i+m}) \;\; = \;\; \frac{4}{(k-1)^2} \sum_{\substack{j_1, j_2 = 1 \\ j_1 < j_2}}^{k-m} E(Y_{j_1}^2 Y_{j_2}^2). \tag{5.74}$$

Now let $h \in \{1, \ldots, k-1-m\}$. It is easy to see that there are exactly $k-m-h$ combinations of $j_1, j_2 \in \{1, 2, \ldots, k-m\}$ for which $j_2 - j_1 = h$. Hence, by the stationarity of $\{Y_t,\, t \in \mathbb{Z}\}$, we obtain

$$E(A_i A_{i+m}) \;=\; \frac{4}{(k-1)^2} \sum_{h=1}^{k-1-m} (k-m-h)\gamma(h). \tag{5.75}$$

The proof is complete. □

Proof of Theorem 5.1.3 According to Lemma 5.1.1 and Slutsky's Theorem, it is sufficient to show that

$$\sqrt{\frac{n}{k}}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V} \;\to\; N(0,\, \tau^2)$$

Using the representation in (5.65), we have

$$\sqrt{\frac{n}{k}}\boldsymbol{V}'\boldsymbol{A}_d\boldsymbol{V} \;=\; \frac{1}{\sqrt{nk}} \sum_{i=(k+1)/2}^{n-(k-1)/2} A_i. \tag{5.76}$$

Furthermore, according to Lemma A.2, the sequence $\{A_i,\, i \in \mathbb{Z}\}$ is strictly stationary and strongly mixing with geometric rate. By Theorem 18.5.3 of Ibragimov and Linnik (1971), it only remains to show that $E|A_i|^{2+\delta/2} < \infty$ and

$$\tau^2 \;=\; k^{-1}\Big(E(A_i^2) + 2\sum_{m=1}^{\infty} E(A_i A_{i+m})\Big). \tag{5.77}$$

For $E|A_i|^{2+\delta/2} < \infty$ to hold, it is sufficient that $E|Y_t|^{4+\delta}$, which immediately follows from the assumption $E|e_t|^{4+\delta} < \infty$. In order to establish (5.77), we use the result (5.67) from Lemma A.2,

which gives us

$$\begin{aligned}
\tau^2 &= \frac{4}{k(k-1)^2}\left(\sum_{h=1}^{k-1}(k-h)\gamma(h) + 2\sum_{m=1}^{k-2}\sum_{h=1}^{k-1-m}(k-m-h)\gamma(h)\right) \\
&= \frac{4}{k(k-1)^2}\left(\sum_{h=1}^{k-1}(k-h)\gamma(h) + 2\sum_{h=1}^{k-2}\sum_{m=1}^{k-h-1}m\gamma(h)\right) \\
&= \frac{4}{k(k-1)^2}\left(\sum_{h=1}^{k-1}(k-h)\gamma(h) + \sum_{h=1}^{k-2}(k-h)(k-h-1)\gamma(h)\right) \\
&= \frac{4}{k(k-1)^2}\sum_{h=1}^{k-1}(k-h)^2\gamma(h). \qquad (5.78)
\end{aligned}$$

Hence, the result of this theorem follows. □

In the remaining part of this section, we establish bootstrap analogues of the previous results. In the following, let $E^*(\cdot)$ and Var$^*(\cdot)$ denote the conditional expectation $E(\cdot|Y_1,\ldots,Y_n)$ and the conditional variance Var$(\cdot|Y_1,\ldots,Y_n)$. We state the following bootstrap analogue of Lemma A.1.

Lemma A.3 *Suppose that H_0 holds. Let $p_1,\ldots,p_l \in \mathbb{N}\cup\{0\}$ and $i_1 < i_2 < \ldots < i_l$. If there exists at least one $s \in \{1,\ldots,l\}$ such that p_s is an odd number, then*

$$E^*\left(Y_{i_1}^{*p_1}\ldots Y_{i_l}^{*p_l}\right) = 0. \qquad (5.79)$$

Proof of Lemma A.3 By the bootstrap construction, e_t^* is symmetrically distributed. Hence, the result can be obtained by the same arguments as in the proof of Lemma A.1. □

In the following we restrict ourselves to GARCH(1,1)processes. The next result is a special case of Lemma 4.1 part (d) of Maercker and Moser (2009). We re-state it here for the later proof of Theorem 5.2.1.

Lemma A.4 *Suppose* $\{Y_t, t \in \mathbb{Z}\}$ *follows a stationary and ergodic GARCH(1,1) process satisfying* $E|Y_t|^{2p} < \infty$ *for some* $p > 2$*, then*

$$E^*\big(Y_{i_1}^{*2}Y_{i_2}^{*2}\big) \quad \xrightarrow{P} \quad E\big(Y_{i_1}^2Y_{i_2}^2\big) \tag{5.80}$$

as $n \to \infty$. The result of Maercker and Moser (2009) is stated more generally for expectations of the form $E^* \prod_{s=1}^4 Y_{i_s}^{*k_s}$ with $k_s \in \{0, 2\}$ for $s = 1, \ldots, 4$. Since we only consider a special case, we can relax their moment condition to $E|e_t|^{4+\delta} < \infty$ for some $\delta > 0$.

Proof of Theorem 5.2.1 It is sufficient to show that

$$T_n^* \quad \to \quad N(0,\, \tau^2)$$

as $n \to \infty$ in probability, with τ^2 as given in Theorem 5.1.2. To establish the asymptotic normality of T_n^*, we use the same techniques applied in the proof of Theorem 5.1.3, but in a bootstrap framework. First, we show that instead of T_n^* we may consider the statistic

$$\sqrt{\frac{n}{k}}\boldsymbol{V^{*\prime}}\boldsymbol{A}_d\boldsymbol{V^*} \quad = \quad \frac{1}{\sqrt{nk}} \sum_{i=(k+1)/2}^{n-(k-1)/2} A_i^* \tag{5.81}$$

where

$$A_i^* \quad = \quad \frac{1}{k-1} \sum_{\substack{j_1, j_2 = i-(k-1)/2 \\ j_1 \neq j_2}}^{i+(k-1)/2} Y_{j_1}^* Y_{j_2}^*. \tag{5.82}$$

The proof is exactly the same as for Lemma 5.1.1, except that we replace any expectation of the form $E\big(Y_{j_1}^2Y_{j_2}^2\big)$ by $E^*\big(Y_{j_1}^{*2}Y_{j_2}^{*2}\big)$ and use the fact that $E^*\big(Y_{j_1}^{*2}Y_{j_2}^{*2}\big)$ converges to $E\big(Y_{j_1}^2Y_{j_2}^2\big)$ in probability as $n \to \infty$.

Next, we derive a statement similar to Lemma A.2. Note that when n is sufficiently large, the estimated GARCH(1,1) parameters $\hat{\alpha}_0$, $\hat{\alpha}_1$ and $\hat{\beta}_1$ are sufficiently close to the true parameters α_0,

α_1 and β_1 so that the bootstrap sequence $\{Y_t^*,\, t \in \mathbb{Z}\}$ is strictly stationary and strongly mixing with geometric rate. Similarly as in the proof of Lemma A.2, we can show that also the sequence $\{A_i^*,\, i \in \mathbb{Z}\}$ is strictly stationary and strongly mixing with geometric rate. The result on the second moments of $\{A_i^*,\, i \in \mathbb{Z}\}$ is again obtained by replacing $E(Y_{j_1}^2 Y_{j_2}^2)$ with $E^*(Y_{j_1}^{*2} Y_{j_2}^{*2})$ and applying Lemma A.4.

Finally, similar to the proof of Theorem 5.1.3, we apply Theorem 18.5.3 of Ibragimov and Linnik (1971) to the bootstrap sequence $\{A_i^*,\, i \in \mathbb{Z}\}$, by which we obtain that $\sqrt{n/k}\boldsymbol{V}^{*\prime}\boldsymbol{A}_d\boldsymbol{V}^*$ conditional on $Y_1, \ldots, Y_n$ converges to a normal distribution with mean 0 and variance τ^2. The proof is complete. $\square$

Chapter 6

Summary and Future Work

In this dissertation, we proposed to apply the *linearization* technique to three different time series modeling and forecasting problems: (1) tracking multiple frequencies; (2) constructing PIs of returns and volatilities of the ARCH/GARCH processes; (3) testing for ARCH/GARCH effects as well as detecting trends in the presence of ARCH/GARCH effect (conditional heteroscedasticity).

In Chapter 2, we introduce a regularized AR (RAR) approximation to estimate the multiple frequencies hidden in a sinusoidal process by the regularized AR (RAR) approximation. The utilization of regularization allows us, in practice, to fit a "longer" AR model and consequently to reduce the bias of the frequency estimates. Our next goal is to complete the proof of Conjecture 2, which theoretically justifies that as the sample size $n \to \infty$ and the order of AR approximation $k \to \infty$ such that $k/n \to 0$, the RAR frequency estimates are strongly consistent. Also, an interesting extension of RAR is to apply banding and thresholding as regularization techniques instead of ridge, as well as to utilize bootstrap-based method to select an "optimal" regularizer. A similar AR approximation can be used to forecast ARFIMA processes based on its AR(∞)

representation.

In Chapter 3, we adopt a sieve bootstrap procedure to construct PIs of returns and volatilities of the ARCH/GARCH processes. The linear AR/ARMA form of the squared ARCH/GARCH process allows us to apply the LS estimation, which considerably reduces the computational costs while preserving coverage and sharpness of PIs. In future work, we will provide asymptotic results on the consistency of the bootstrap PIs. Also, our bootstrap procedure is not truly "sieve" according to Bühlmann's definition (Bühlmann, 1997), though it is linear. We assume that the order of ARCH/GARCH model is given instead of approximating the observed process by an AR model whose order is selected by some information criterion, such as AIC or BIC. In practice, the true order of the underlying process is typically unknown, so we will extend the proposed approach by sieve approximation and investigate its properties. As noted in the existing sieve bootstrap literature (Bühlmann, 1997, 2000; Alonso et al., 2003, 2004), AIC or BIC may not be the optimal model order selection criteria. We are also interested in proposing an information criterion which is both consistent and efficient for sieve bootstrap. As an alternative to sieve bootstrap, we will further apply blockwise bootstrap (Künsch, 1989; Liu and Singh, 1992) to construct PIs of returns and volatilities and justify their asymptotic properties.

A similar linear bootstrap scheme can also be used for diagnostic testing. In Chapter 4, we propose a robust Lagrange Multiplier (LM) test for detecting ARCH/GRACH effects, which utilizes the permutation/bootstrap procedure to obtain critical values. We show that the permutation LM (PLM) test is exact while the bootstrap LM (BLM) test is asymptotically correct. Our numerical studies indicate that the proposed re-sampling algorithms significantly improve size and power of the LM test in both skewed and heavy-tailed processes. In future work, we will compare PLM and BLM to other model misspecification tests, e.g., portmanteau and goodness-of-fit tests (Mcleod and Li, 1983; Wong and Li, 1995 and 2002; Ling and Li, 1997). Also, we plan to

extend the permutation/bootstrap scheme to sensitivity analysis, model diagnostics as well as the assessment of the predictive distribution of future returns and volatilities.

In Chapter 5, we suggest a nonparametric trend (a deterministic function of time) test in the presence of ARCH/GARCH effects based on the heteroscedastic ANOVA, so-called NT-GARCH. The proposed test can effectively detect non-monotonic trend under ARCH/GARCH effects, especially when data are seasonal. We will complete the proof of Conjecture 2 on the asymptotic normality of the NT-GARCH test statistic, as $n \to \infty$ and $k_n \to \infty$ such that $k_n/n \to 0$. Moreover, we plan to further investigate the applications of the linear form of ARCH/GARCH model in change-point and regime switching detection, as well as the construction of prediction intervals for trend.

Beyond this dissertation, our more general topic is *linearization* with nonparametric methods in time series analysis. The philosophy behind the proposed dissertation is to use simple linear methods with least restrictive assumptions in time series modeling and forecasting. *Linearization* is a broad research topic and can be applied to tackle many other time series problems, such as the modeling and forecasting of various extensions of the ARCH/GARCH model, long memory processes and nonparametric models. With the help of the fast progressing computing facilities, there is a large potential of developing different nonparametric techniques to incorporate with the *linearization* approaches. Our long term goal is to introduce a set of *linearization* methodology in different aspects of time series analysis with purely data-driven estimation and prediction methods, which is preferable to their complicated nonlinear counterparts.

Bibliography

[1] Abraham, B. and Ledolter, J. 2005. *Statistical Methods for Forecasting*. Wiley, New York.

[2] Akaike, H. 1969. Fitting autoregressive models for prediction. *Annals of The Institute of Statistical Mathematics* **21**, 243-247.

[3] Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**(6): 716-723.

[4] Akritas, M.G. and Papadatos, N. 2004. Heteroscedastic one-way ANOVA and lack of fit test. *Journal of the American Statistical Association* **99**: 368-382.

[5] Alonso, A.M., Pẽna, D. and Romo, J. 2002. Forecasting time series with sieve bootstrap. *Journal of Statistical Planning and Inference* **100**: 1-11.

[6] Alonso, A.M., Pẽna, D. and Romo, J. 2003. On sieve bootstrap prediction intervals. *Statistics and Probability Letters* **65**: 13-20.

[7] An, H.Z., Chen Z. G. and Hannan, E.J. 1982. Autocorrelation, autoregression and autoregressive approximation. *Annals of Statistics* **10**(3): 926-936.

[8] Andersen, T.G. and Bollerslev, T. 1998. Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39**: 885-905.

[9] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. 2001. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* **96**: 42-55.

[10] Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. 2003. Modelling and forecasting volatility. *Econometrica* **71**: 579-625.

[11] Anderson, R. 2008. *Modern Methods for Robust Regression*. SAGE Publications.

[12] Artis, M., Hoffmann, M., Nachane, D. and Toro, J. 2004. The detection of hidden periodicities: A comparison of alternative methods. *EUI working paper ECO 10*.

[13] Artis, M., Clavel, J., Hoffmann, M. and Nachane, D. 2007. Analyzing strongly periodic series in the frequency domain: A comparison of alternative approaches with applications. *CEPR Discussion Papers 6517*.

[14] Ahdesmäki, M., Lähdesmäki, H., Pearson, R., Huttunen, H. and Yli-Harja, O. 2005. Robust detection of periodic time series measured from biological systems. *BMC Bioinformatics* **6**(117).

[15] Baillie, R.T. and Bollerslev, T. 1992. Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics* **52**: 91-113.

[16] Baillie, R.T., Bollerslev, T. and Mikkelsen, H. O. 1996. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **74**: 3-30.

[17] Barabanov, Ye., and Gel, Y. 2005. Strong consistency of the least-squares method with a polynomial regularizer for infinite AR models. *Automation and Remote Control* **1**(66): 92-108.

[18] Bera, A.K. and Higgins, M. L. 1993. ARCH models: properties, estimation and testing. *Journal of Economic Surveys* **7**: 305-366.

[19] Berk, K.N. 1974. Consistent autoregressive spectral estimates. *The Annals of Statistics* **2**(3): 489-502.

[20] Berkes, I., Horváth, L. and Kokoszka, P. 2003. GARCH processes: structure and estimation. *Bernoulli* **9**(2): 201-227.

[21] Bickel, P. and Bühlmann, P. 1999. A new mixing notion and functional central limit theorems for a sieve bootstrap in time series. *Bernoulli* **5**(3): 413-466.

[22] Bickel, P.J. and Freedman, D. 1981. Some asymptotic theory for bootstrap. *Annals of Statistics* **9**(6): 1196-1217.

[23] Bickel P. and Gel, Y. 2011. Banded regularization of covariance matrices in application to parameter estimation and forecasting of time series. *To appear in the Journal of the Royal Statistical Society, Series B*.

[24] Bickel, P.J. and Levina, E. 2008a. Regularized estimation of large covariance matrices. *Annals of Statistics* **36**(1): 199-227.

[25] Bickel, P.J. and Levina, E. 2008b. Covariance regularization by thresholding. *Annals of Statistics* **36**(6): 2577-2604.

[26] Bickel, P.J. and Sakov, A. 2008. On the choice of m in the m out of n bootstrap and its application to confidence bounds for extreme percentiles. *Statistica Sinica* **18**: 967-985.

[27] Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**: 307-327.

[28] Bollerslev, T. 2008. Glossary to ARCH (GARCH). CREATES Research Paper: 2008-2049.

[29] Bose, A. 1988. Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics* **16**: 1709-1722.

[30] Bougerol, P. and Picard, N. 1992a. Strict stationarity of generalized autoregressive processes. *Annals of Probability* **20**: 1714-1730.

[31] Bougerol, P. and Picard, N. 1992b. Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* **52**: 115-127.

[32] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 2008. *Time Series Analysis: Forecasting and Control*. Wiley, New York.

[33] Breusch, T.S. and Pagan, A.R. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica* **47**(5): 1287-1294.

[34] Brillinger, D.R. 1987. Fitting cosines: some procedures and some physical examples. *in: B. MacNeill, G.J. Umphrey, Applied Probability, Stochastic Processes, and Sampling Theory, Ed.I*. Reidel, Boston. 75-100.

[35] Brockwell, P.J. and Davis, R.A. 1991. *Time Series: Theory and Methods*. Springer, New York.

[36] Bühlmann, P. 1995. Moving-average representation of autoregressive approxiamation. *Stochastic Processes and their Applications* **60**: 331-342.

[37] Bühlmann, P. 1997. Sieve bootstrap for time series. *Bernoulli* **3**: 123-148.

[38] Bühlmann, P. 2002. Bootstraps for time series. *Statistical Science* **17**(1): 52-72.

[39] Bunea, F., Wegkamp, M. H., and Auguste, A. 2006. Consistent variable selection in high dimensional regression via multiple testing. *Journal of Statistical Planning and Inference* **136**(12): 4349-4364.

[40] Burn, D.H. 1994. Hydrologic effects of climatic change in West Central Canada. *Journal of Hydrology* **160**: 53-70.

[41] Burn, D.H. and Hag Elnur, M.A., 2002. Detection of hydrologic trends and variability. *Journal of Hydrology* **255**: 107-122.

[42] Cao, R., Febrero-Bande, M., González, W., Prada-Sánchez, J.M. and García-Jurado, I. 1997. Saving computer time in constructing consistent bootstrap prediction intervals for autoregressive processes. *Communications in Statistics - Simulation and Computation* **26**: 961-978.

[43] Chen, C.S, Li, C.H. and Su, H.C. 2008. A nonlinear time series analysis using two-stage genetic algorithms for streamflow forecasting. *Hydrological processes* **22**: 3697-3711.

[44] Chen, Z. 1988. An alternative consistent procedure for detecting hidden frequencies. *Journal of Time Series Analysis* **21**(2): 301-317.

[45] Chen, B. and Gel, Y.R. 2010. Autoregressive frequency detection using regularized least squares. *Journal of Multivariate Analysis* **101**(7): 1712-1727.

[46] Chen, B. and Gel, Y.R. 2011. Regularized autoregressive multiple frequency estimation. *Journal of the Iranian Statistical Society*. *Accepted*.

[47] Chen, B. and Gel, Y.R. 2011. Robust lagrange multiplier test for detecting ARCH/GARCH effect using permutation and bootstrap. *Under revision*.

[48] Chen, B., Gel, Y.R., Balakrishna, N. and Abraham, B. 2011. Computationally efficient bootstrap prediction intervals for returns and volatilities in ARCH and GARCH processes. *Journal of Forecasting* **30**: 51-71.

[49] Chen, B., Gel, Y.R., Abraham, B. and El-Shaarawi. A. 2011. An ANOVA-type nonparametric trend test under ARCH/ GARCH effect. *Working paper*.

[50] Chen, Z., Wu, K. and Dahlhaus, R. 2000. Hidden frequency estimation with Data Tapers. *Journal of Time Series Analysis* **21**(2): 113-142.

[51] Conrad, C. and Karanasos, M. 2006. The impulse response function of the long memory GARCH process. *Economic Letters* **90**: 34-41.

[52] Cook, R.D. and Weisberg, S. 1983. Diagnostics for heteroscedasticity in regression. *Biometrika* **70**: 1-10.

[53] Darania, P., Abadian, E., Oskoi, A.V. 2006. Linearization method for solving nonlinear integral equations, *Mathematical Problems in Engineering*: 1-10.

[54] Demos, A. and Sentana, E. 1998. Testing for GARCH effects: a one-sided approach. *Journal of Econometrics* **86**: 97-127.

[55] Doukhan, P. 1994. *Mixing Properties and Examples: Lecture Notes in Statistics Vol 85*. Springer, New York.

[56] Duan, Z., Pardo, B. and Zhang, C. 2010. Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions. *IEEE Transactions on Audio, Speech, and Language Processing* **18**(8): 2121-2133.

[57] Dufour, J.M., Khalaf, L., Brenard, J.T. and Genest, I. 2004. Simulation based finite sample tests for heteroskedasticity and ARCH effects. *Journal of Econometrics* **122**(2): 317-347.

[58] EI-Shaarawi, A.I. and Niculescu, S.P. 1992. On Kendall's Tau as a test of trend in time series data. *Environmetrics* **3**: 389-411.

[59] Elasmi-Ksibi, R., Besbes, H. and López-Valcarce, R. 2010. Frequency estimation of real-valued single-tone in colored noise using multiple autocorrelation lags. *Signal Processing* **90**(7): 2303-2307.

[60] Engle, R.F. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987-1007.

[61] Engle, R.F., Hendry, D.F. and Trumble, D. 1985. Small-sample properties of ARCH estimators and tests. *Canadian Journal of Economics* **18**: 66-93.

[62] Engle, R.F. and Bollerslev, T. 1986. Modeling the persistence of conditional variances. *Econometric Reviews* **5**: 1-50.

[63] Engle, R.F. and Ng, V. K. 1993. Measuring and testing the impact of news on volatility. *Journal of Finance* **48**(5): 1749-1778.

[64] Engle, R.F. and Patton, A.J. 2001. What good is a volatility model? *Quantitative Finance* **1**: 237-245.

[65] Engle, R.F. 2002. Dynamic conditional correlation: A simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* **20**: 339-350.

[66] Engle, R.F. and Rangel, J.G. 2008. The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *The review of financial studies* **21**(3): 1187-1222.

[67] Fan, J. and Yao, Q. 2003. *Nonlinear Time Series*. Springer, New York.

[68] Fomin, V.N. 1998. *Optimal filtering Voloum I: Filtering of stochastic processes*. Kluwer Academic Publishers, Dordrecht.

[69] Francq, C. and Zakoïan, J.M. 1998. Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* **83**: 369-394.

[70] Francq, C. and Zakoïan, J.M. 2000. Covariance matrix estimation for estimators of mixing weak ARMA models. *Journal of Statistical Planning and Inference* **68**: 145-165.

[71] Furrer, R. and Bengtsson, T. 2007. Estimation of high-dimensional prior and posteriori covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis* **98**(2): 227-255.

[72] Gately, E. 1996. *Neural Networks for Financial Forecasting*. John Wiley and Sons, New York.

[73] Gel, Y.R. and Fomin, V.N. 2001. Identification of an unstable ARMA equation. *Mathematical Problems in Engineering* **7**: 97-112.

[74] Gel, Y.R. and Barabanov, A. 2007. Strong consistency of the regularized least-squares estimates of infinite autoregressive models. *Journal of Statistical Planning and Inference* **137**: 1260-1277.

[75] Gel, Y.R., Miao, W. and Gastwirth, J.L. 2007. Robust directed tests of normality against heavy tailed alternatives. *Computational Statistics and Data Analysis* **51**: 2734-2746.

[76] Gel, Y.R. and Gastwirth, J.L. 2008. The robust Jarque-Bera test of normality. *Economics Letters* **99**(1): 30-32.

[77] Giraitis, L., Kokoszka, P.S. and Leipus, R. 2000. Stationary ARCH models: dependence structure and central limit theorem. *Econometric Theory* **16**: 3-22.

[78] Godfrey, L.G. 1979. Testing for multiplicative heteroskedasticity. *Journal of Econometrics* **8**: 227-236.

[79] Godfrey, L.G., Orme, C.D. and Santos Silva, J.M.C. 2004. Simulation based tests for heteroskedasticity: some further results. *University of Manchester, Economic studies, Discussion paper 0417*.

[80] Gradshteyn, I.S., Ryzhik, I.M., Jeffrey, A. and Zwillinger, D. 2000. *Table of integrals, series, and products*. Academic Press, 6th edition.

[81] Granger, C.W.J. and Joyeux R. 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* **1**: 15-29.

[82] Gregory, A.W. 1989. A nonparametric test for autoregressive conditional heteroscedasticity: a Markov-Chain approach. *Journal of Business and Economic Statistics* **7**: 107-115.

[83] Grenander, U. 1981. *Abstract Inference*. Wiley, New York.

[84] Häedle, W., Horowitz, J.L. and Kreiss, J.P. 2003. Bootstrap methods for time series. *International Statistical Review* **71**: 435-459.

[85] Hamed, K.H. and Rao, A.R. 1998. A modified Mann-Kendall trend test for autocorrelated data. *Journal of Hydrology* **204**: 182-196.

[86] Hamed, K.H. 2008. Trend detection in hydrologic data: The Mann-Kendall trend test under the scaling hypothesis. *Journal of Hydrology* **349**: 350-363.

[87] Hamed, K.H. 2009. Exact distribution of the Mann-Kendall trend test statistic for persistent data. *Journal of Hydrology* **365**: 86-94.

[88] Hannan, E.J. 1971. Nonlinear time series regression. *Journal of Applied Probability* **8**: 767-780.

[89] Hannan, E.J. 1978. The asymptotic theory of linear time series models. *Journal of Applied Probability* **10**(1):130-145.

[90] Hannan, E.J. and Huang, D. 1993. On-Line Frequency Estimation. *Journal of Time Series Analysis* **14**(2): 147-161.

[91] Hipel, K.W. and MeLeod, A.I. and Weiler, R.R. 1988. Data analysis of water quality time series in Lake Erie. *Water Resources Bulletin* **24**(3): 533-544.

[92] Hockey, T.A. 1999. *Galileo's Planet: Observing Jupiter Before Photography*. Institute of Physics Publishing, London.

[93] Hoeffding, W. 1952. The large-sample power of tests based on permutation of observations. *The Annals of Mathematical Statistics* **30**: 420-447.

[94] Hoerl, A. E. and Kennard, R. W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**(1): 55-67.

[95] Hofmann, G and Balakrishnan, N. 2006. A nonparametric test for trend based on initial ranks. *Journal of Statistical Computation and Simulation* **76**(9): 829-837.

[96] Hofmann, T. Schölkopf, B. and Smola, A.J. 2008. Kernal methods in machine learning. *Annals of Statistics* **36**(3): 1171-1220.

[97] Horn, R.A. and Johnson, C.R. 1994. *Topics in Matrix Analysis*. Cambridge, England: Cambridge University Press.

[98] Horváth, L and Kokoszka, P. 2001. Large sample distribution of weighted sums of ARCH(p) squared residuals correlations. *Econometric Theory* **17**: 283-295

[99] Hosking, J.R.M. 1981. Fractional differencing. *Biometrika* **68**: 165-176.

[100] Houdré, C. and Kedem, J. 1995. A note on autocovariance estimation in the presence of discrete spectra. *Statistics and Probability Letters* **24**: 1-8.

[101] Huang, D. 1993. On low and high frequency estimation. *Journal of Time Series Analysis* **17**(4): 351-365.

[102] Hurvich, C.M. and Tsai, C.L. 1989. Regression and time series model selection in small samples. *Biometrika* **76**: 297-307.

[103] Ibragimov I.A. and Linnik Yu.V. 1971. *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.

[104] Ing, C.K. and Wei, C.Z. 2003. On same-realization prediction in infinite-order autoregressive process. *Journal of Multivariate Analysis* **85**: 130-155.

[105] Ing, C.K. and Wei, C.Z. 2005. Order selection for same-realization predictions in autoregressive processes. *Annals of Statistics* **33**(5): 2423-2474.

[106] Jarque, C.M., Bera, A.K. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* **6**(3): 255-259.

[107] Kedem, B. 1994. *Time series analysis by higher-order zero crossing*. IEEE Press, Piscataway, New Jersey.

[108] Kendall, M.G., Gibbons, J.D., 1990. Rank Correlation Methods, 5th edition, Griffin, London.

[109] Khalil, H.K. 2002. *Nonlinear Systems*. Prentice Hall, New Jersey.

[110] Koenker, R. 1981. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics* **17**: 107-112.

[111] Kreiss, J. P. 1988. Asymptotic statistical inference for a class of stochastic processes. *Habilitationsschrift, Universität Hamburg*.

[112] Kreiss, J.P. and Franke, J. 1992. Bootstrapping stationary autoregressive moving average models. *Journal of Time Series Analysis* **13**: 297-317.

[113] Kristensen, D. and Linton, O. 2006. A closed-form estimator for the GARCH (1,1) model. *Econometric Theory* **22**: 323-337.

[114] Künsch, H.R. 1989. The jackknife and the bootstrap for general stationary observations. *Annals of Statistics* **17**: 1217-1241.

[115] Kuwabara, M. and Watanabe, N. 2006. A fuzzy trend model for long-term financial time series and its identification. *Fuzzy Information Processing Society, 2006. NAFIPS 2006. Annual meeting of the North American*. 478-483.

[116] Lahiri, S.N. 2003. *Resampling Methods for Dependent Data*. Springer, New York.

[117] Lau, S.S., Sherman, P.J. and White, L.B. 2002. Asymptotic statistical properties of AR spectral estimators for processes with mixed spectra. *IEEE Transactions on Information Theory* **48**(4): 909-917.

[118] Ledoit, O. and Wolf, M. 2004. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* **88**(2): 365-411.

[119] Lee, J.H.H. 1991. A Lagrange multiplier test for GARCH models. *Economics Letters* **37**: 265-271.

[120] Léger, C. and Romano, J. 1990. Bootstrap adaptive estimation: the trimmed mean example. *Canadian Journal of Statistics* **18**: 297-314.

[121] Lehmann, E.L and Romano, J.P. 2005. *Testing Statistical Hypotheses*. Springer, New York.

[122] Li, T.H., Kedem, B. and Yakowitz, S. 1994. Asymptotic normality of sample autocovariances with an applications in frequency estimation. *Stochastic Processes and their Applications* **52**: 329-349.

[123] Li, T.H. and Kedem, B. 1998. Tracking abrupt frequency changes. *Journal of Time Series Analysis* **19**: 69-82.

[124] Lindner, A.M. 2009. Stationarity, mixing, distributional properties and moments of GARCH(p, q) processes. *Handbook of financial time series* Part 1: 43-69.

[125] Ling, S. and Li, W.K. 1997. Diagnostic checking of nonlinear multivariate time series with multivariate ARCH errors. *Journal of Time Series Analysis* **18**(5): 447-464.

[126] Liu, Y., Nie, Z., Zhao, Z and Liu, Q. 2011. Generalization of iterative Fourier interpolation algorithms for single frequency estimation. *Digital Signal Processing* **21**(1): 141-149.

[127] Liu, R. and Singh, K. 1992. Moving blocks jackknife and bootstrap capture weak dependence. In R. LePage and L. Billard (eds), *Exploring the Limits of Bootstrap*. New York: Wiley.

[128] Maercker, G. and Moser, M. 2009. Yule-Walker type etimators in GARCH(1,1) model: asymptotic normality and bootstrap. *Working paper*.

[129] MacKisack, M.S. and Poskitt, D.S. 1989. Autoregressive frequency estimation. *Biometrika* **76**(3): 565-575.

[130] MacKisack, M.S. and Poskitt, D.S. 1990. Some properties of autoregressive estimates for processes with mixed spectra. *Journal of Time Series Analysis* **11**(4): 325-337.

[131] Mann, H.B. 1945. Nonparametric tests against trend. *Econometrica* **13**: 245-259.

[132] McLeod, A.I. and Li, W.K. 1983. Diagnostic checking ARMA time series models using squared-residual autocorrelations. *Journal of Time Series Analysis* **4**(4): 269-273.

[133] Medeiros, M.C. and Veiga, A. 2009. Modeling multiple regimes in financial volatility with a flexible coefficient GARCH(1,1) Model. *Econometric Theory* **25**(1): 117-161.

[134] Meinshausen, N. 2005. Lasso with relaxation. *UC Berkeley technique report*.

[135] Miguel, J.A., Olave, P. 1999. Bootstrapping forecast intervals in ARCH models. *Test* **8**(2): 345-364.

[136] Moffatt, M. 2008. About.com State-Space Approach. *Economics Glossary*.

[137] Nelson, D.B. 1990. Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**: 318-344.

[138] Nelson, D.B. 1991. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* **59**: 347-370.

[139] Noguchi, K., Gel, Y. and Duguay, C.R. Bootstrap-based tests for trends in hydrological time series, with application to ice phenology data. *Submitted*.

[140] Park, B.J. 2002. An outlier robust GARCH model and forecasting volatility of exchange rate returns. *Journal of Forecasting* **21**: 381-393.

[141] Parzen, E. 1962. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* **33**: 1065-1076.

[142] Parzen, E. 1974. Some recent advances in time series modeling. *IEEE Transactions on Automatic Control* **19**: 723-730.

[143] Pascual, L., Romo, J. and Ruiz, E. 2004. Bootstrap predictive inference for ARIMA processes. *Journal of Time Series Analysis* **25**: 449-465.

[144] Pascual, L., Romo, J. and Ruiz, E. 2006. Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics and Data Analysis* **50**: 2293-2312.

[145] Peguin-Feissolle, A. 1999. A Comparison of the Power of Some Tests for Conditional Heteroscedasticity. *Economics Letters* **63**: 5-17.

[146] Pianosi, F. and Raso, L. 2008. Flood forecasting for heteroscedastic streamflow processes. *Proceedings of the 17th World Congress The International Federation of Automatic Control Seoul, Korea*.

[147] Pisarenko, V.F. 1973. The retrieval of harmonic from a covariance function. *Geophysical Journal International* **33**(3): 347-366.

[148] Politis, D.N. 2003. The impact of bootstrap methods on time series analysis. *Statistical Science* **28**: 219-230.

[149] Politis, D.N., 2007. Can the stock market be linearized? *Economics working paper: Department of Economics, Univerisity of California at San Diego*.

[150] Poskitt, D.S. 2006. Autoregressive approximation in nonstandard situations: the fractionally integrated and non-invertible cases. *Annals of the Institute of Statistical Mathematics* **59**(4): 697-725.

[151] Poon, S.H. 2005. *A Practical Guide to Forecasting Financial Market Volatility*. John Wiely and Sons, Chichester, UK.

[152] Prony, G.R. 1795. Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastiques et sur celles de la force expansive de la vapeur de leau et de la vapeur de lalkool, à. différentes températures. *Journal de L'École Polytechnique Floréal et. Plairial, an III* voloum 1, cahier 22, 24-76.

[153] Quinn, B.G. 2000. On Kay's Frequency Estimator. *Journal of Time Series Analysis* **21**(6): 701-712.

[154] Quinn, B.G. and Hannan, E.J. 2001. *The Estimation and Tracking of Frequency*. Cambridge University Press, UK.

[155] Rao, C.R. and Zhao, L.C. 1993. Asymptotic behavior of maximum likelihood estimates of superimposed exponential signals. *IEEE Transactions on Signal Processing* **42**: 1461-1464.

[156] Raunig, B. 2008. Detecting ARCH effects in non-gaussian time series. *Journal of Financial Econometrics* **6**(2): 271-289.

[157] Ray, B.K. 1993. Modeling long-memory processes for optimal long-range prediction. *Journal of Time Series Analysis* **14**: 511-525.

[158] Ray, B. and Crato, N. 1996. Model Selection and Forecasting of Long-range Dependent Processes. *Journal of Forecasting* **15**: 107-125.

[159] Reeves, J.J. 2005. Bootstrap prediction intervals for ARCH models. *International Journal of Forecasting* **21**: 237-248.

[160] Romano, J.P. 1989. Bootstrap and randomization tests of some nonparametric hypotheses. *The Annals of Statistics* **17**(1): 141-159.

[161] Romano, J.P. and Thombs, L.A. 1996. Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* **91**(434): 590-600.

[162] Rosenblatt, M. 1952. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* **23**: 470472.

[163] Rosenblatt, M. 1956. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics* **27**: 832-837.

[164] Sakai, H. 1984. Statistical analysis of Pisarenkos method for sinusoidal frequency estimation. *IEEE Transactions on Acoustics, Speech and Signal Processing* **32**: 95-101.

[165] Serfling, R.J. 1980. *Approximation theorem of mathematical statistics*. Wiley, New York.

[166] Schölkopf, B., Burges, C.J. and Smola, A.J. 1998. *Advances in kernel methods: support vector learning*. MIT press.

[167] Schroeder, D.J. 1999. *Astronomical Optics*. Academic Press, New York.

[168] Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* **6**(2): 461-464.

[169] Shephard, N. 1996. Statistical aspects of ARCH and stochastic volatility models. In *Time Series Models in Econometrics, Finance and Other Fields*, Cox, D.R., Hinkley, D.V. and Barndorff-Nielsen, O.E. (eds). Chapman and Hall, London. 1-67.

[170] Shibata, R. 1980. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* **8**: 147164.

[171] Song, K.S and Li, T.H. 2006. On convergence and bias correction of a joint estimation algorithm for multiple sinusoidal frequencies. *Journal of the American Statistical Association* **101**(474): 830-842.

[172] Stoica, P., Friedlander, B. and Söderström, T. 1987. Asymptotic bias of the high-order autoregressive estimates of sinusoidal frequencies. *Circuits Systems Signal Process* **3**(6): 287-298.

[173] Stoica, P., Söderström, T. and Ti, F. 1989. Asymptotic properties of the high-order Yule-Walker estimates of sinusoidal frequencies. *IEEE Transactions on Acoustics, Speech and Signal Processing* **37**(11): 1721-1734.

[174] Sugihara, G. and May, R.M. 1990. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* **344**: 734-741.

[175] Taylor, S.J. 2005. *Asset Price Dynamics and Prediction*. Princeton University Press, Princeton, New Jersey.

[176] Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B* **58**(1): 267-288.

[177] Tibshirani, R. and Taylor, J. 2010. The solution path of the generalized lasso. *Annals of Statistics* **39**(3): 1335-1371.

[178] Thombs, L.A. and Schucany, W.R. 1990. Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association* **85**: 486-492.

[179] Tsay, R.S. 2002. *Analysis of Financial Time Series*. Wiley Interscience, New York.

[180] Tufts, D.W. and Kumaresan, R. 1982. Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood. *Proceedings of the IEEE* **70**: 975-989.

[181] Tychonoff, A. N. 1943. On the stability of inverse problems. *Doklady Akademii Nauk SSSR39*: 195-198.

[182] Walker, A.M. 1971. On the estimation of a harmonic component in a time series with stationary independent residuals. *Biometrika* **58**: 21-36.

[183] Wang, P. 2006. Stochasticity, nonlinearity and forecasting of streamflow processes. ISO Press.

[184] Wang, W., Van Gelder, P.H.A.J.M., Vrijling, J.K. and Ma, J. 2005. Testing and modelling autoregressive conditional heteroscedasticity of streamflow processes. *Nonlinear Processes in Geophysics* **12**: 55-66.

[185] Wang, L. and Akritas, M.G. 2006. Two-way heteroscedastic ANOVA when the number of levels is large. *Statistica Sinica* **16**: 1387-1408.

[186] Wang, L. and Van Keilegom, I. 2007. Nonparametric test for the form of parametric regression with time series errors. *Statistica Sinica* **17**: 369-386.

[187] Wang, L., Akritas, M.G. and Van Keilegom, I. 2008. An ANOVA-type nonparametric diagnostic test for heteroscedastic regression models. *Journal of Nonparametric Statistics* **20**(5): 365-382.

[188] Weiss, A.A. 1986. Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* **20**: 107-131.

[189] West, K.D. and Cho, D. 1995. The predictive ability of several models of exchange rate volatility. *Journal of Econometrics* **69**: 367-391.

[190] Withers, C.S. 1981. Central limit theorems for dependent variables I. *Probability Theory and Related Fields* **57**4: 509-534.

[191] Wong, H. and Li, W.K. 1995. Portmanteau test for conditional heteroscedasticity, using ranks of squared residuals. *Journal of Applied Statistics* **22**(1): 121-134.

[192] Wong, H. and Li, W.K. 2002. Detecting and diagnostic checking multivariate conditional heteroscedastic time series models. *Annals of The Institute of Statistical Mathematics* **54**(1): 45-59.

[193] Wu, W. B. and Pourahmadi, M. 2003. Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika* **90**(4): 831-844.

[194] Yule, G.U. 1921. On the time correlation problem. *Journal of the Royal Statistical Society* **84**: 497-510.